

Greece should abandon a short-sighted policy

The Greek government's research funding originated in an era of support from the European Union that is coming to an end. The country's potential deserves a much greater focus on fundamental research.

Following the example set by the United States, nearly all nations of the European Union (EU) now accept that a sensible investment in basic research is required for the national economic good. Greek politicians, on the other hand, remain unshakably convinced of the opposite. If they have their way, as from this year there will be no single source of money for research that does not require an industrial partner. Greece is poor, they argue, and cannot afford the luxury of funding research for pleasure rather than business.

A battle is raging between the General Secretariat of Research and Technology (GSRT), a department of the Greek Ministry of Development, and the 500 or so tenured researchers who work in its 44 institutes. The researchers argue that the government's approach is short-sighted, and warn that forcing academics to forge artificial partnerships with the still-embryonic Greek industry, which generally has no interest in research, will result in money simply being misspent.

They are right to be incensed. Last year the GSRT research institutes were evaluated by international panels of experts which concluded that some scattered research excellence can be identified. But this is despite, not because of, government policies, which keep salaries uncompetitively low, provide insufficient money to keep institutes in good repair and stocked with basic instrumentation, and, most importantly, provide no source of competitive national research funds.

The best researchers have survived on EU Framework programme grants and other external sources of money, and on EU structural funds which Greece, as one of the poorer EU countries, receives to build up its economic base. But there will almost certainly be no grants available in the next Framework programme, which starts in 2005, at a

time when structural funds will also end. The next programme will probably concentrate its research money in large centres of excellence.

A handful of GSRT institutes could compete for such funding, but the lifeblood for the strong research groups in others will be cut off. And the government is making no contingency plans. On the contrary, it has proposed adding a requirement for an industrial partner to qualify for the only source of basic research project money, which exists courtesy of structural funds. And it is in the process of amending its law on research institutes to bring them closer to government policy.

In no other EU country are research and industrial policy so tightly coupled, and so misconceived, as in Greece. The past decade's attempts to force researchers to kick-start the Greek economy have had no obvious effect on its industry's research and development activities. Intensifying the concentration of money spent on forcing this issue is wrong. The money would be better spent investing in basic research, which will deliver when it has industry to deliver to: when the Greek economy has found its feet by other means. Moreover, Greece's total expenditure on research is, at 0.5% of its gross domestic product (GDP), by far the lowest in the European Union — lower even than that of several former communist countries, such as Poland (0.76% GDP) and Hungary (0.73%), which can by no means be described as 'richer' than Greece.

As external funds dwindle, Greece must prioritize increasing its research expenditure, and create a national research policy that includes a system of competitive funds for basic research. Its government cannot claim that it cannot afford to do so, for it self-evidently cannot afford not to. ■

Post-genomic cultures

Like it or not, big biology is here to stay.

From anatomy to physiology to molecular biology, biological research has for decades been on an ascending slope of increasing systems complexity. The availability of genome sequences heralds a sharp acceleration in the upturn of that slope. Today biology is dominated by small, investigator-driven groups of 'wet' biologists. But techniques looking at the integrated functioning of hundreds to thousands of genes and proteins — such as microarray and other 'whole genome' analyses — inevitably raise the question of whether the study of one or a few genes or proteins is sophisticated enough, given the need to understand cells, pathways and whole organisms.

Wet biology will continue to be essential, but deep understanding of biological processes will also require a more quantitative molecular biology, with a firm grounding in the behaviour of complex networks. A shift towards an information-oriented systems view of biology, which grasps both mathematically and biologically the many elements of a system, and the relationships among them that allow the construction of an organism, is under way.

But the social change required to make this shift painlessly should not be underestimated. Research groups and institutions will need to take bold steps to embrace computational biology and high-throughput technologies, and to move the emphasis from investigator-driven grant proposals to supporting large multidisciplinary teams that extend beyond departmental and institutional boundaries.

These issues have been considered by some, for example by the National Institutes of Health's new Biomedical Information Science and Technology Initiative (<http://grants.nih.gov/grants/bistic/bistic.cfm>). The Bio-X Program at Stanford University is also one of several new flagship interdisciplinary programmes tackling this 'bigger picture of biology'. But one cannot help feeling that the message has yet to reach the bulk of the troops and, moreover, that few of the generals are providing the leadership for them to attack these new fronts confidently. "What does the human genome sequence mean for me, my research and my institution?" is a question all biologists should be asking themselves. ■





Fond farewell
Outgoing energy
minister bolsters
morale at US labs
p548



Healthy future
Canada sets its sights
high for medical
research
p549



Bank statement
Colombian institute
paralysed by debt
collectors
p550



Rice squad
Transgenic rice
spearheads fight
against malnutrition
p551

Hughes institute will put down roots to develop research tools

Colin Macilwain, Washington

In a radical change of direction, the Howard Hughes Medical Institute (HHMI) will this week announce plans to build a \$500 million research campus near Washington DC.

The creation of a permanent research base is a new move for Hughes, the largest medical research charity in the United States. Until now, it has concentrated on supporting an élite of individual investigators based at US universities.

The campus will be built on a green-field site near Washington Dulles Airport in northern Virginia. HHMI officials envisage that collaborations there between biologists, chemists, physicists and engineers will specialize in developing new research tools. "It is at the interface between the disciplines that the sparks are going to fly," says Tom Cech, HHMI's president. The campus is expected to open its doors by 2005.

The Howard Hughes Investigator programme — which supports 348 researchers to the tune of around \$1 million a year each — has worked well, most observers say. Its researchers include many of the



Watch this space: Cech (right) wants to keep the new campus 'fresh'.

United States' most eminent life scientists. And the absence of any permanent research infrastructure has shielded Hughes from criticisms of empire-building or from any sense of stagnation.

But Hughes needs to expand — it is

required by law to spend 3.5% of its fast-growing endowment on medical research each year. And biomedical researchers' growing appetite for sophisticated and expensive research tools has convinced HHMI officials that there is an opportunity for a vibrant new centre specializing in the development and use of such tools.

"Ramping up the number of principal investigators [to 348] was the right strategy for the 1990s," says Cech. "But biology is now switching over, to benefit from bigger installations."

HHMI officials mention Pat Brown, a Hughes investigator and geneticist at Stanford University, as the sort of researcher who might be induced to join the new establishment. Brown has pioneered the development of microarray techniques in genetics. David Clayton, the vice-president for scientific development at HHMI, says it will attract "people who really enjoy bringing state-of-the-art technology to bear" on scientific problems.

As well as working on research tools, the



Ethics watchdog to oversee drugs trials in Third World

Matthew Davis, Washington

Tighter controls for US-funded medical research involving human subjects in the developing world were announced last week by the US Department of Health and Human Services.

In response to growing concerns — from both the press and scientific advisory groups — over ethical standards for clinical trials in developing countries, the department has set up an Office of International Activities.

The new division, to be located within the Office for Human Research Protections (OHRP), will oversee, but not regulate, federally funded trials in developing countries and provide ethical guidance.

The OHRP already regulates federally funded scientists working abroad, although research by drug companies is controlled by

the Food and Drug Administration.

A recent series of articles in *The Washington Post*, criticizing the way US researchers operate in developing countries, depicted large pharmaceutical companies — and some federally funded researchers — as increasingly exploiting vulnerable patient populations in poor countries.

For example, Pfizer was criticized for a 1996 clinical trial in Nigeria, in which its antibiotic trovafloxacin was tested on children and infants with meningitis. *The Washington Post* portrayed the company as taking advantage of a meningitis epidemic, but Pfizer rejected the charges, saying that the trial played "an important role in investigating a potential breakthrough in oral therapy for this terrible epidemic".

The Washington Post articles have prompted calls for a congressional health

panel to seek hearings on the issue, and "a legislative response".

The issue is also being driven by the National Bioethics Advisory Commission and the World Medical Association, which had intensified their involvement before the newspaper series began.

The organizations were spurred into action partly by complaints that began three years ago over a Ugandan clinical trial in which HIV-positive pregnant women were given placebos even though, critics claimed, a proven therapy was available that could have saved the unborn child from infection.

The two organizations advise scientists not to use placebos in developing countries when effective treatments exist. But critics argue that without a placebo group it is often impossible to gauge the effectiveness of a new therapy.

campus will serve as what Gerry Rubin, Hughes vice-president for research, calls a "research hotel", where teams of visiting scholars, who may or may not be Hughes investigators, can spend time together working on projects of interest. The groups could come "for a few months, or even a couple of years", says Rubin. "No university can hold space open for that."

The facility will be built on a recently acquired 281-acre site. It will house around 24 principal investigators and up to 300 support staff, plus visiting scientists. Hughes has put aside \$500 million for the project over 10 years, including planning, construction and \$50 million a year for operations after 2005. A director will be appointed about a year before operations begin.

Hughes officials say that investigators at the campus will not have tenure, although their five-year appointments will be renewable. "We've spent a lot of time discussing how to keep it fresh," says Cech. "We think we can keep people moving in and out of the place."

Cech and Rubin arrived at HHMI at the start of 2000 (see *Nature* 402, 334–335; 1999) and have since been working on a new direction for the charity. The announcement is its central component. "This is something that the new leadership group here has cooked up," says Cech. ■

Weapons labs escape FBI action

Irwin Goodwin, Washington

Bill Richardson, the US energy secretary in Bill Clinton's administration, spent his last two days in office trying to restore the morale and motivation of thousands of scientists at US nuclear weapons labs.

He announced that, after a seven-month investigation, the FBI will not be bringing criminal charges against any workers at the Los Alamos National Laboratory (LANL) in New Mexico, where two computer hard drives containing secret nuclear-weapons data temporarily went missing last spring.

The FBI "was unable to determine responsibility for the disappearance ... and found no evidence that the classified information contained on the hard drives had been compromised", the energy department said in a statement.

At least four Los Alamos employees had been investigated in connection with the disappearance. Coming close on the heels of the high-profile Wen Ho Lee affair — in which a Taiwan-born computer scientist working at LANL had been accused of passing classified information to China — the investigation had seriously shaken the morale of the laboratory's scientists.

Richardson also extended the energy department's contract with the University of California to manage LANL and its sister



Morale booster: Richardson's final acts.

Lawrence Livermore laboratory near San Francisco until September 2005. The university has always managed the weapons labs, but had recently been criticized by Congress and the energy department for allowing apparent security lapses.

In addition, Richardson withdrew the threat of polygraph tests for workers, which he had announced after Wen Ho Lee was sacked nearly two years ago. Initially he wanted 8,000 employees, mostly scientists, to take the test, but reduced this to 400 after many resigned and young researchers refused to accept jobs. "We have great scientists in the weapons labs," Richardson says, "and we have to make sure we can retain and recruit the best and brightest." ■

Journal will publish accused scientist's work

Rex Dalton, San Diego

After months of debate, a leading journal has decided it will consider for publication the research by a renowned geochemist accused of criminal offences unconnected to his work.

Geochimica et Cosmochimica Acta has been in turmoil for months after two articles were submitted co-authored by Yale University professor Antonio Lasaga, who has been charged with child sex crimes. Lasaga specializes in the kinetics of geochemical processes.

The Reed Elsevier journal is jointly sponsored by the Geochemical Society and the Meteoritical Society. The journal's editor, Frank Podosek, a professor at Washington University in St Louis, Missouri, decided not to publish research bearing Lasaga's name, saying he wanted to protect the societies and the journal. "I didn't want to create the impression we were looking the other way," Podosek says.

Two other co-authors on the articles — Hiroshi Ohmoto of Pennsylvania State University and Andreas Lutge of Rice University in Texas — immediately filed complaints with the journal's publications committee.



Lutge, an associate professor now seeking tenure at Rice after studying under a German fellowship with Lasaga at Yale, is upset by the publication delay. "I have to feed my family," says Lutge, "and proceed with my career. I spent four years at Yale — I can't throw out that work."

Lasaga and his attorneys declined interview requests. Lasaga has pleaded guilty to two federal felonies for possession of child pornography, but has since attempted to withdraw the pleas. He also faces state charges for alleged sexual abuse of a young boy. He has pleaded innocent in this case.

As tensions mounted in the geochemical societies over whether to publish the papers, Podosek polled the journal's 55 associate editors on their positions. About two-thirds of those responding supported publication of Lasaga's work, says Podosek. But Podosek stood firm in his view.

A compromise was finally worked out by Geochemical Society president Michael Hochella of Virginia Polytechnical Institute,

and the University of Arizona's Michael Drake, the former president of the Meteoritical Society.

Under the arrangement, Podosek will not handle research articles where Lasaga is an author — an associate editor will direct them through the normal peer-review process.

"This has been an extraordinarily difficult situation for all concerned," say Hochella and Drake in a joint statement. "Our societies do not have a mechanism to deal with reviewing papers based on anything other than scientific merit and the standard ethical aspects of scientific research."

At their annual meetings later this year, the societies' boards will consider adopting a policy to "define such a mechanism", the statement says. Meanwhile, Podosek and Lasaga's co-authors were satisfied with the outcome. "They made the right decision," says Ohmoto.

But others were more critical. Nicholas Cozzarelli, a molecular biologist at the University of California and editor of the *Proceedings of the National Academy of Sciences*, says: "The [Lasaga] papers should be treated like any others. It is a slippery slope when editors make decisions on non-scientific matters." ■

Canada pours funds into health research

David Spurgeon, Montreal

The newly created Canadian Institutes of Health Research (CIHR) has signalled in its first round of funding that it will keep its promise to support “innovative interdisciplinary ways to conduct research”.

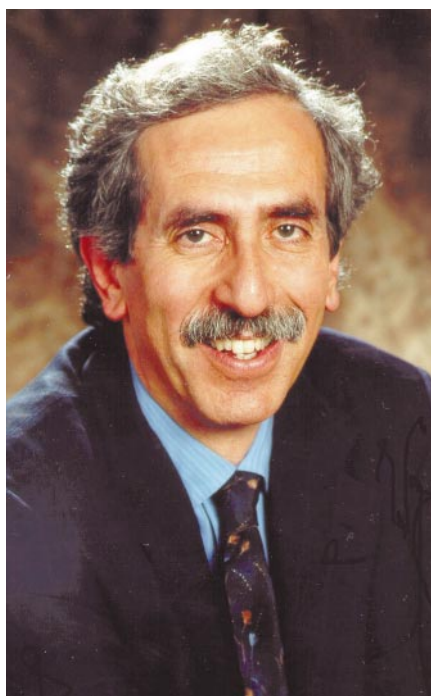
The agency announced last week that it will distribute Can\$64 million (US\$43 million) across nearly 500 diverse research projects, involving some 600 researchers in 100 different institutions.

The CIHR's budget over the next three years is effectively twice that of the Medical Research Council of Canada, which the CIHR replaced last June (*Nature* **405**, 722; 2000) — and the new agency considerably broadens the scope of health research in Canada.

Many of the projects are on a scale that Canadian researchers have not previously enjoyed. Nineteen Interdisciplinary Health Research Teams (IHRTs), comprising more than 500 investigators in 91 institutes across Canada and internationally, will share Can\$15.5 million.

These teams will undertake research projects on such topics as the process of ageing; opiate addiction; breast-cancer susceptibility; and a genetic, epidemiological and population-based approach to the impact and control of colorectal cancer.

A further 19 large-scale multidisciplinary projects have been funded under the Community Alliances for Health Research (CAHR) programme, which aims to foster excellence in research of relevance to com-



Stand and deliver: awards will make Canada ‘the place to be’ for health research, says Bernstein.

munity groups and agencies in biomedical and clinical research, and in health services generally. These projects will include research into community genetics, women's unpaid care-giving, chronic illness in rural healthcare systems, and marine and coastal workplace health and safety.

Funds were also approved for clinical

trials, equipment and maintenance grants. Pilot projects in the national Genomics Research Program include a project at the University of Toronto to develop a rapid ‘gene-trapping’ method. Research grants are also being awarded under the federal government's HIV/AIDS strategy and the Health Canada/CIHR research initiative on hepatitis C.

Robert S. Bell, an IHRT recipient at Mount Sinai Hospital, Toronto, says the programme has given himself and his colleagues “a unique opportunity” to bring together the three major centres in Canada for musculoskeletal cancers, and to combine clinical information with basic biology.

Before the CIHR was set up, says Bell, “there was no agency that would allow us to bring together patient information, tumour specimens, basic science information, clinical information and clinical outcomes and put that together into a big database that lets us to go from the patient to the laboratory and back again.”

In a message to Canadian researchers, CIHR president Alan Bernstein said the awards demonstrate the CIHR's commitment to transform research in Canada.

“If we establish the right balance between investigator-initiated and strategic research, along with the right process to develop the right initiatives,” said Bernstein, “I am convinced that we will be able to create [an] environment that will make Canada ‘the place to be’ for health research in the twenty-first century.” ■

Poor coordination 'wastes US research into global change'

Mark Schrope, Washington

The infrastructure for coordinating US research into global change must be redesigned if politicians are to make properly informed decisions on big issues such as greenhouse-gas emission policies. This is the conclusion of a report from the National Research Council (NRC) which aims to influence the new Bush administration.

The report — *The Science of Regional and Global Change: Putting Knowledge to Work* — says poor coordination within and between agencies such as the space agency NASA and the National Oceanic and Atmospheric Administration (NOAA) has stopped politicians and researchers making full use of results.

For example, NASA's ocean remote-sensing data must be calibrated and validated using measurements taken at the ocean surface which might best be carried out by NOAA. But although such work is critical for integrating the NASA data into models of global change, NOAA might see this as a low priority.

Moreover, the report says that critical challenges — such as accurate predictions on a global scale — call for improved computing capabilities and global monitoring systems which are beyond the capacity of individual agencies.

Charles Kennel, chairman of the NRC's Committee on Global Change Research, which produced the report, says accurate predictions of changes on a regional scale are also crucial.

"People can't answer the most important question in environmental science — what's going to happen to me?" he says. "The present system is not really capable of doing that."

The report proposes that global-change research should be coordinated through a new, centralized government body. This would filter results to policymakers and fund critical projects.

The US Global Change Research Program, set up in 1990 to coordinate the efforts of government and other organizations, could be upgraded to take care of these new tasks, says the report. Or a new body could be created.

"I very much agree that some high-level entity is needed, because we now have an amalgam of separate agency programmes where the goal of each agency is very different," says Marvin Geller, president of the atmospheric sciences section of the American Geophysical Union.

Bank raids malaria centre in dispute over landlord's debt

Xavier Bosch

Research at the Immunology Institute of Bogotá (IIB) in Colombia — headed by controversial biochemist Manuel Patarroyo — has been paralysed by debt collectors.

In a move that has precipitated angry demonstrations, a Spanish bank impounded the institute's equipment because it is owed \$2.5 million by the hospital in which the institute is located.

Three weeks ago, the BBVA Banco Ganadero bank, which has extensive interests in Latin America, took control of the IIB's supercomputers, DNA sequencers and nuclear magnetic resonance equipment, which is used to determine the three-dimensional structures of proteins.

The bank accepts that the IIB is only associated with, and not owned by, the Foundation Hospital San Juan de Dios, which owes the money. But the bank's president José María Ayala, a Spaniard, says that, after several attempts to redeem debts from the hospital by confiscating property, there was "nothing else left of value to impound". He says he had recommended that Patarroyo leave the hospital and move elsewhere.

But Patarroyo says the bank had no right to involve the IIB, which, he argues, is almost entirely independent of the foundation that runs the hospital, being directly funded by the Colombian presidency. Only the building itself and half the salaries of 20 out of the 168 researchers are provided by the foundation, he says.

The foundation, which manages four health and research centres in Bogotá on an annual budget of \$60 million, was driven into severe debt by Colombia's general financial difficulties combined with the privatization of the foundation.

The bank's action has generated intense local ill-feeling — Patarroyo says he has had to persuade an angry demonstration of



Patarroyo: work on malaria vaccine has been paralysed by removal of institute equipment.

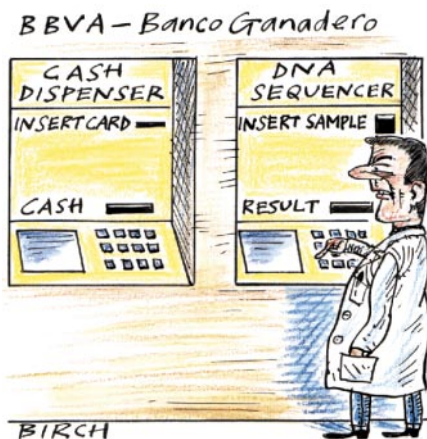
students and young researchers not to storm the bank. And Colombian newspapers have published letters calling for investors to withdraw their capital from the bank.

"There is even a bit of anti-Spanish feeling among people now," says Patarroyo, who has strong research links with Spain.

Patarroyo became famous in Colombia when he created a synthetic malaria vaccine containing peptides derived from three asexual blood-stage proteins of the protozoan parasite *Plasmodium falciparum*.

But he was frequently attacked by the scientific community during the late 1980s and early 1990s over methodological issues, including the launch of what was considered by many to be a badly designed clinical trial involving over 20,000 Colombians. The vaccine was eventually shown to have low but significant efficacy, which his research team has since been trying to improve by incorporating more peptides. Patarroyo says he hopes to have an improved vaccine ready for clinical testing within two years.

The IIB received about \$3 million from the Colombian presidency last year, under an agreement linking funding to research output. Ayala says he hopes to find a solution that would allow him to have the impounded goods returned, perhaps through the intervention of the Colombian presidency.





Golden future: from right, Peter Beyer, Ingo Potrykus and colleagues with the rice that they developed.

Designer rice to combat diet deficiencies makes its debut

Quirin Schiermeier, Munich

Heralding the launch of the humanitarian 'Golden Rice' project, the first free samples of this new type of rice — genetically engineered to contain vitamin A precursors — were last week shipped to the International Rice Research Institute (IRRI) in Los Baños, Philippines.

This is the first of many planned deliveries of the rice to non-commercial research institutes in China, India, Africa and Latin America. The institutes will conduct bio-safety studies, and then use traditional breeding techniques to confer the beneficial traits of the transgenic rice to strains adapted to local growing conditions.

Thanks to a unique arrangement between the two inventors of Golden Rice and the agribusiness Syngenta, to which the inventors assigned all commercial rights, subsistence farmers in any developing country can cultivate Golden Rice varieties, once available, licence-free. Subsistence farmers are defined as having annual earnings below \$10,000 per year.

The inventors are plant scientist Ingo Potrykus, who retired from the Swiss Federal Institute of Technology in Zurich in 1999, and biochemist Peter Beyer of the University of Freiburg in Germany.

The Golden Rice project was promoted by Potrykus, who wanted his research to help combat the vitamin A deficiencies prevalent in many poor countries, particularly those relying on rice as a major food source. Rice plants do not normally produce carotenoids, vitamin A precursors, in the grain. A 'humanitarian board' made up of the two

inventors and representatives of the Rockefeller Foundation, the World Health Organization and the biotechnology industry, will oversee the distribution of the rice to the research institutes.

Syngenta — the world's largest agribusiness, which this week announced the completion of the entire rice genome (see right) — and several other companies holding patents relating to different steps in the development of Golden Rice, have agreed to waive licence fees for the project. "This was very complicated to arrange, because we had used 70 patents from 32 companies and universities," says Potrykus.

Nor was it straightforward for Potrykus and his colleagues to donate their own intellectual property rights. They initially found it hard to convince the European Commission, which funded some of the work, that donating licences for humanitarian purposes would not interfere with the commission's requirement that its research should strengthen the competitiveness of European industry. But eventually the commission welcomed the Golden Rice achievement.

Others are more cynical about Golden Rice, seeing it as a means for the plant biotechnology industry to polish its tarnished image. Environmental groups such as Greenpeace have called the project a Trojan horse, saying it opens the door to extended cultivation of genetically modified crops in the developing world.

"This is not true," says Potrykus. "We are making sure that varieties important to the poor will be used, not fashionable varieties for the urban middle class."

Commercial sector scores success with whole rice genome

**David Dickson, London
and David Cyranoski, Tokyo**

Syngenta, the world's largest agribusiness corporation, and gene-discovery company Myriad Genetics have sequenced the complete rice genome. But the sequence will be available only through contracts and will not be published.

The genome is the second largest to have been sequenced to date (smaller only than the human genome) and the first for a crop plant. The project took only 18 months and has produced a genome map that the companies say is "99.5% complete". The publicly funded effort to sequence the rice genome's 430 million DNA bases, led by Japan, is three years away from completion.

Myriad, which is based in Salt Lake City, says that its proprietary high-throughput sequencing facilities enabled it to complete the genome six months ahead of schedule — earning a \$3 million bonus from Syngenta as part of a \$33 million deal.

The companies' success has raised worries about the impact on poorer countries that rely on rice as a staple food crop. But Syngenta says it will provide information and technology that can produce crop improvements for subsistence farm-related research "without royalties or technology fees".

The company says that it will make its rice genome data — derived from the Nipponbare variety — available to researchers through "research contracts".

Given the apparent — and apparently unexpected — similarity between the rice genome and those of various cereals, the rice sequence is expected to aid the study of crops such as wheat, corn and barley.

Syngenta's announcement has shaken partners in the publicly funded International Rice Genome Sequencing Project (IRGSP), which has sequenced 5% of the genome to completion and an additional 25% to the level of quality achieved by Syngenta's sequence.

Takuji Sasaki, principal investigator of the IRGSP, says the project's partners will meet next week and discuss Syngenta's results. He says the claim of 99.5% completeness will be difficult to verify "as the company won't show its cards". But he worries that politicians may nonetheless consider the job done and pull the plug on the publicly funded project. This would be unwise, he says, as the IRGSP's sequence will be both more accurate than Syngenta's and freely available.

Earthquake tragedy in India spares nuclear power station

New Delhi India's most severe earthquake in 50 years killed tens of thousands and left many more homeless in Gujarat last week. But the region's atomic power station and space installations were left unscathed.

Ahmedabad, the state's second largest city, is home to the Space Applications Centre, which builds all the payloads for Indian satellites — including the one scheduled to be launched in February — and the Physical Research Laboratory, which does basic studies in space science. "Our structures did not develop even a crack, although high-rise buildings in the immediate neighbourhood collapsed," a spokesman for the space department told *Nature*.

Authorities said that the atomic power station at Kakrapar 150 km southeast of Ahmedabad also suffered no damage. On the one-to-five scale into which India has been divided, Ahmedabad and Kakrapar are in seismic zone three, whereas Bhuj, which was at the epicentre of last week's quake, is in zone five.

Along with the tragedy, the Bhuj quake has raised some difficult questions for seismologists. The area is in the northern

fringe of the stable continental region, and last experienced such a powerful quake in 1819. The recurrence cycle for a quake of this size in this area is expected to be a few thousand years.

Bill Gates provides funding boost to AIDS vaccine

Geneva The prospect of developing an affordable AIDS vaccine received a boost last weekend with the news that the charitable foundation set up by Microsoft founder Bill Gates is donating \$100 million to the International AIDS Vaccine Initiative (IAVI).

The latest gift from the Bill and Melinda Gates Foundation, announced at the World Economic Forum in Davos, Switzerland, means that the IAVI has now secured \$230 million. It needs \$550 million to meet its goal of propelling at least six vaccine candidates into efficacy trials by 2007. The Internet company Yahoo! also pledged a further \$5 million to the initiative.

"This is a marker for the future," says Andrew McMichael of the University of Oxford, who is working on a vaccine with backing from the IAVI. McMichael's vaccine, being developed with colleagues in Kenya, consists of a DNA vaccine plus the same sequences expressed in a weakened *Vaccinia* virus.

Cassini sends back snaps of Jupiter

Washington Forget Stanley Kubrick — this is how Jupiter and its volcanic moon Io really looked at the dawn of 2001, as photographed by the outbound Cassini spacecraft on 1 January. Along with a boost in velocity, the Jupiter fly-by gave the \$3.4 billion spacecraft a chance to show its capabilities before arriving at Saturn in 2004 for a four-year tour.



NASA/JPL/UNIVERSITY OF ARIZONA

Cassini's cameras, spectrometers, radio antennas and other instruments have not disappointed. By the time the Jupiter photography session ends in March, the spacecraft will have returned 25,000 pictures of the planet, most of them focused on its stormy atmosphere. Even though it flew past at a relatively distant 9.7 million kilometres, Cassini's time-lapse pictures will be invaluable to researchers studying the dynamics of Jupiter's weather. The onboard cameras, which improve on those of the Galileo spacecraft in terms of image size and sharpness, also resolved the small moon Himalia for the first time. And a new kind of sensor, called the magnetospheric imaging instrument, took synthetic pictures of the gigantic field of energetic particles surrounding the planet.

NASA head breaks records in Bush administration

Washington NASA's charismatic administrator, Dan Goldin, who was appointed by George Bush the elder nearly nine years ago, has been asked by Vice-President Dick Cheney to stay on in the job temporarily, until a successor can be found. This gives Goldin the unique distinction of being the first NASA chief in the agency's 43-year history to serve in three different administrations.

Anonymous donors back biotech research

Washington It has been a week of anonymous philanthropy. The Johns Hopkins School of Medicine in Baltimore has received \$58.5 million to launch a cell engineering institute. This donation could provide an alternative to federal funding for stem-cell research. Although the US National Institutes of Health last year said it would support stem-cell research projects, President George W. Bush's office has sent signals that he may reverse this decision. The Hopkins Institute for Cell Engineering will focus on selecting, modifying and reprogramming human cells as potential therapeutic transplants.

Meanwhile, in New York, the Rensselaer Polytechnic Institute has said it will build a new biotechnology facility with \$70 million of a recent anonymous \$130 million gift, one of the largest ever given to a US academic institution.

Geneticists unstick Tokyo's favourite snack

Tokyo Researchers at Japan's Applied Bacteriology Laboratory, part of the National Food Research Institute, have found the gene that makes the fermented soybeans known as natto sticky. Natto is a staple part of the diet



for many in the greater Tokyo area — but it usually offends the uninitiated with both its texture and pungent smell.

The researchers randomly disrupted genes in *Bacillus subtilis*, the fermentation bacterium that makes the polyglutamate which coats the beans with their sticky film. They identified a regulatory gene that is essential for polyglutamate production.

Along with possible medical applications related to polyglutamate's calcium-absorbing properties, the researchers think it might someday be possible to adjust natto stickiness to specific tastes.

Notice of erratum In response to many queries, this is to give notice that a formal erratum will appear next week concerning the paper "The protein-protein interaction map of *Helicobacter pylori*" (*Nature* **409**, 211–215; 2001). This paper conforms to our policy of data access: Supplementary Information on *Nature's* website contains the set of protein interactions reported and analysed in the paper (see <http://www.nature.com>). Owing to a proof-reading oversight in the *Nature* office, this statement and link were not included in the published paper.



Taming Africa's killer lake

An international team of scientists and engineers is in Cameroon to begin 'degassing' Lake Nyos, scene of a 1986 natural disaster in which a cloud of carbon dioxide killed more than 1,700 people. Tom Clarke assesses the risks and benefits.

When death came, it was silent and sudden. On the night of 21 August 1986, a cold and misty blanket of carbon dioxide gas flowed through the valleys that run down from Lake Nyos in northwest Cameroon. Hugging the ground, the dense, dank cloud suffocated livestock and smothered more than 1,700 people, many as they slept.

The region surrounding Lake Nyos is volcanic, and so sulphurous volcanic gases were initially suspected as the cause of the disaster. But subsequent investigations by teams of scientists dispatched to the remote highland lake suggested that as much as 80 million cubic metres of CO₂, formerly held in solution in the lake's depths, had been released explosively at its surface. Such an event, called a limnic eruption, had been recorded only once before — at nearby Lake Monoun, 35 kilometres from Lake Nyos. That eruption, which killed 37 people in 1984, was kept secret at first, as the Cameroonian authorities

had suspected terrorist involvement.

Expeditions to Lake Nyos since the 1986 disaster have allowed researchers to unravel the geological processes that underlie the lake's lethal tendencies¹⁻³. Armed with this information, a small group of scientists and engineers — funded to the tune of US\$450,000 by the Office of US Foreign Disaster Assistance, with additional help from the French and Cameroonian governments — arrived in the area in January to attempt to 'degas' the lake, and so prevent it killing again. The plan is simple — the limited budget and the lack of decent roads into the region mean it has to be. The team will lower a length of high-density polyethylene pipe deep into the lake and pump CO₂-rich water towards the surface, where the gas can bubble out in a controlled way.

Most experts agree that degassing the lake is a good idea. But some say there is a small chance that the effort could trigger another limnic eruption and put lives at risk.



In one night, 1,700 people and many livestock were suffocated by CO₂ from Lake Nyos. A 1995 expedition has since tested degassing equipment.

Although villages surrounding Lake Nyos have been evacuated to avoid a repeat of the 1986 catastrophe, farmers and their livestock still venture into the area, lured by its lush pastures.

Under pressure

The water in Lake Nyos is divided into two layers. The surface layer, around 50 metres deep, is unremarkable, and is fed by rainfall and streams. But the lake's depths, which extend to around 200 metres, contain large quantities of dissolved CO₂. This became obvious when scientists began sampling the water after the 1986 eruption. "As we brought the samples to the surface the bottles exploded," says Sam Freeth, director of the Geological Hazards Research Unit at the University of Wales in Swansea, who visited the region two weeks after the disaster.

Further investigations revealed that the lake is fed from beneath by soda springs, which continually pump CO₂-rich water

into its depths. The pressure of the overlying water keeps the gas in solution. "It's like a bottle of champagne," says Alain Bernard, a geochemist at the Free University of Brussels. "Before you open it, you don't see any bubbles."

The thin boundary layer that separates the surface and deep waters is "like a membrane or barrier that's relatively strong and stable", says Bernard. But on 21 August 1986, this balance was disturbed, perhaps by a mudslide, rainfall, or unusual patterns of wind blowing across the lake — no one knows for sure. "Something stirred it up," says Freeth. The result was that water from the depths rushed to the surface and vast quantities of CO₂ came foaming out.

Since the eruption, CO₂ has been building up again. Based on estimates of how much gas was released in the disaster and measurements of the flow of CO₂ into the lake's depths, scientists think that, by 1992, the concentration of CO₂ had reached the same level as immediately before the 1986 event. Since then, the rise in CO₂ levels seems to have tailed off somewhat. "But everyone would agree the lake is as dangerous now as it was before the disaster," says Freeth.

Soda streams

Many of the scientists and engineers behind the current project visited the lake shortly after the disaster and have maintained a fascination with the problem of making it safe. Their long-term goal is to reverse the build-up of CO₂. As a start, they are assembling and deploying a 200-metre-long pipe, lowered from a platform in the middle of the lake (see diagram, right). If this test is successful, more pipes may be added later on.

As pumps draw water up the pipe, the CO₂ will begin to come out of solution. As it does so, the gas and water will start to rise up the pipe under their own buoyancy, sucking more water from the depths and creating a soda fountain that should become self-sustaining. The basic concept was suggested soon after the disaster⁴⁻⁶ and was proven using a prototype during a previous expedition to Lake Nyos in 1995.

A second platform will also be floated on the lake, housing instruments to monitor the water's flow rate, its temperature, and the concentration of CO₂, which will be inferred from measurements of salinity and electrical conductivity. The instruments will allow the team to see how successful the fountain is in removing CO₂ from the lake. "We hope that it will be sufficient to lower the levels," says Bernard.

Constant monitoring will also be important for safety reasons. The expedition team intends to stay at the lake until 8 February, after which a satellite link will relay data back to the laboratory of one of the project's leaders, Michel Halbwachs at Savoie University in Chambéry, France.

Although the principle of degassing has been proven, there is still a lot to learn about how best to operate such a scheme. The researchers have opted for just one pipe in the first instance, because they are concerned that more might destabilize the lake. "The outcome of this experiment is not to empty the lake of CO₂ but to measure the influence that one pipe has on the stratification of the lake," says Jean-Christophe Sabroux, an engineer with France's Atomic Energy Commission. If it looks as if the degassing is in danger of getting out of control, the pipe is fitted with a valve that can be closed off remotely. The decision on whether to shut it will lie with Cameroonian officials. "They are the ones who will ultimately be affected," says Sabroux.

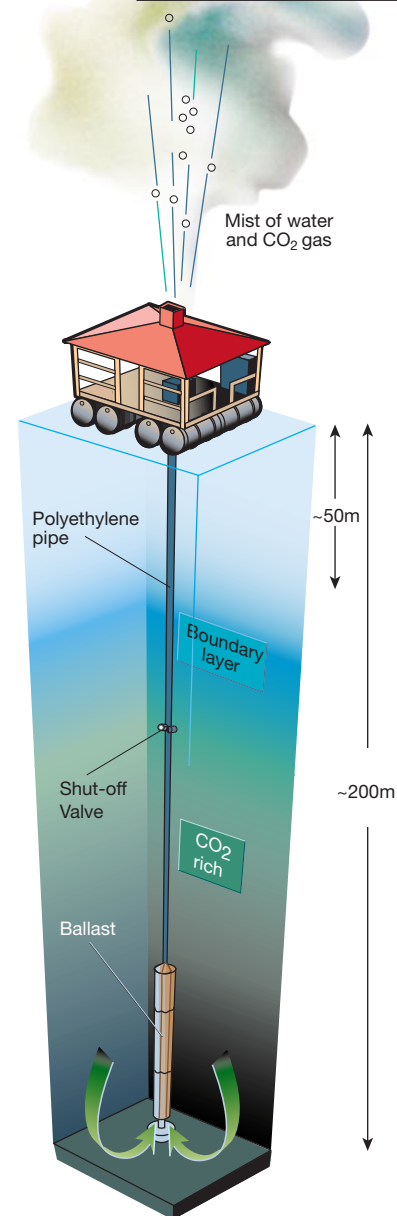
Lakeside manners

Despite these measures, Freeth, who is not involved in the present project, remains worried about the risks. "There's a lot of gas in that lake and I'm concerned that sticking pipes in and bringing up water could destabilize the lake," he says. Freeth fears that CO₂-rich water spouting from the pipe and landing on the lake's surface could sink back down and disturb the boundary layer.

Sabroux plays down the dangers. Calculations made during previous visits to the lake, including the 1995 trial run, suggest that Freeth's concerns are unfounded, he says. By starting cautiously with a single pipe and monitoring it closely, Sabroux says, the team will learn how to handle the killer lake. He argues that scientists cannot simply stand by and let Lake Nyos continue to pose a hazard. "Everyone connected with the project knows they are taking risks, but they are worth facing," he says.

Sabroux adds that there are sufficient safety procedures to protect the expedition team and local people helping with the project. Roads into the area will be closed and a system of CO₂-monitoring stations with alarms has been installed around Lakes Nyos and Monoun. If the alarms sound, all personnel working at Lake Nyos have been instructed to head to areas of high ground where, because CO₂ is more dense than air and tends to hug low ground, they will be safe. Those working on the lake itself have been equipped with breathing apparatus provided by the Cameroonian military.

Even if the degassing project proves a resounding success, the danger posed by Lake Nyos may not be over. One section of the shore, where a waterfall drains the lake during the wet season, is a natural dam made up of loosely packed volcanic rock. The wall of the dam is already heavily eroded⁷, and geologists who have studied the site are concerned that it could collapse. "Perhaps in the not so far distant future, that dam could fail," says Freeth.



Gas station: the 200-metre pipe should allow CO₂ to be extracted from Lake Nyos safely.

If so, according to some predictions, the upper 40 metres of water in the lake could flow down valleys to the north, causing flash floods in heavily populated regions of Nigeria some 150 kilometres away. And if that were to happen before the lake is degassed, it would also almost certainly cause an encore of the 1986 eruption. ■

Tom Clarke is a science writer based in London.

1. Kling, G. W. *et al.* *Science* **236**, 169–175 (1987).
2. Sigurdsson, H. *et al.* *J. Volcanol. Geotherm. Res.* **31**, 1–16 (1987).
3. Nojiru, Y. *et al.* *Nature* **346**, 322–323 (1990).
4. Freeth, S. J. *et al.* *Nature* **348**, 201 (1990).
5. Kling, G. W., Evans, W. C., Tuttle, M. L. & Tanyileke, G. *Nature* **368**, 405–406 (1994).
6. Halbwachs, M., Grangeon, J., Sabroux, J.-C. & Villevalle, A. *C. R. Acad. Sci.* **316**, 483–489 (1993).
7. Lockwood, J. P., Costa, J. E., Tuttle, M. L., Nni, J. & Tebor, S. G. *Bull. Volcanol.* **50**, 340–349 (1988).

♦ <http://perso.wanadoo.fr/mhalb/nyos>

♦ <http://www.biology.lsa.umich.edu/~gwk/research/nyos.html>

Digital history

Some historians of science are moving away from the traditional image of lone scholars poring over ancient manuscripts. Alison Abbott talks to one of history's digital pioneers.



BRIDGEMAN ART LIBRARY

The informatics revolution, already in the process of reshaping biology, is now homing in on the history of science. An ambitious new initiative launched by the Max Planck Institute for the History of Science in Berlin aims to use the cutting edge of information technology to understand how scientific knowledge developed.

The initiative is the brainchild of Jürgen Renn, a physicist-turned-historian and one of the Berlin institute's directors. His idea is that all possible sources of information that have come down over the millennia should be digitized and made available in a form that allows the data to be mined beyond traditional barriers of language and academic discipline.

The basic idea of digitizing historical documents so that they can be accessed by scholars from all over the world on the Internet is not new. Renn, for instance, collaborates in the Cuneiform Digital Library Initiative, which provides virtual access to texts and translations from ancient Mesopotamia, dispersed in museums and private collections. But Renn is one of a new breed of historians who want to extend the approach. They intend to add multiple levels of annotation and to structure their databases so that they become fully searchable, regardless of the

language of origin — not just through key words, but also through the comparison of underlying concepts.

The effort required is vast because — as with genome databases — adding annotation to help interpret data entries cannot be fully automated; it is labour intensive and requires specialist knowledge. Realizing Renn's goals will also require the development of advanced search and matching tools that can analyse texts for contextual relationships between the terms they contain (see *Nature* **405**, 112–115; 2000). More challenging, it will also require the development of translation programs vastly superior to those available today. But the approach should offer historians fresh opportunities to make connections that would otherwise remain obscure.

Privileged access

Ultimately, says Renn, the digital approach should strip history of its esotericism, opening up avenues of study that currently are closed to all but a few individuals with the necessary linguistic skills, expert knowledge or privileged access to documents and artefacts. Once the infrastructure he aims to build is in place, for instance, a historian should no longer need to employ a philolo-

gist to help interpret a manuscript written in an ancient Greek dialect, or in old Italian.

As a proof of principle, with colleagues from Tufts University in Massachusetts, Harvard University and the University of Missouri at Kansas City, Renn is embarking on a project on the history of mechanics, called Archimedes. Shortly before Christmas, it won an initial three-year, US\$1 million grant, funded jointly by the US National Science Foundation (NSF) and the DFG, Germany's main research grant-giving body, as part of the International Digital Libraries Initiative, a project set up and funded by the NSF and Britain's Joint Information Systems Committee.

The project is "both modest and ambitious", says Renn. "Mechanics is a small part of science as a whole, yet the material that could be made available is almost inconceivably vast." He estimates that the project could encompass millions of pages covering almost two millennia and several cultures and languages. Language analysis and translation programs developed for Tufts University's Perseus Project, which provides a digital library of ancient texts with modern translations, will be exploited and extended.

As its starting point, the Archimedes project will digitize and annotate the earliest text



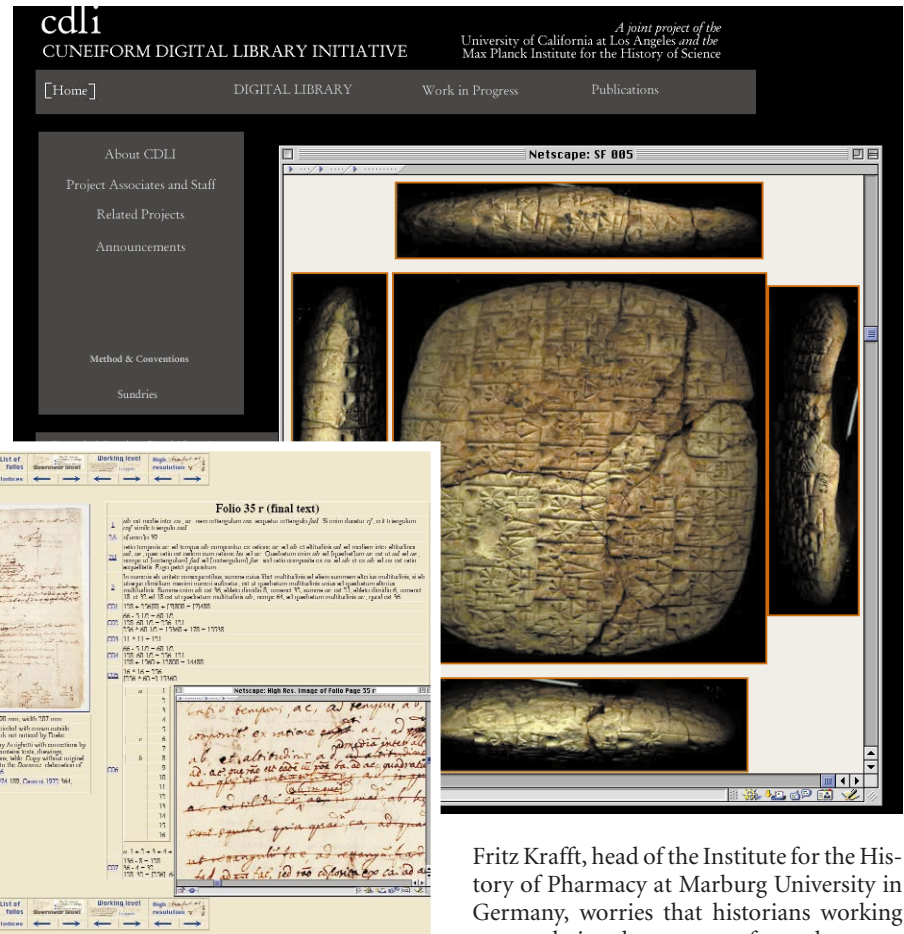
Building to last: Galileo's lab notes, Mesopotamian texts and Florence cathedral's (left) construction archives are among the treasures being put online by Jürgen Renn (above) and his colleagues.

on mechanics, a tract by Aristotle. Other documents and sources of information will be selected by half a dozen core historians, along with 20 or so experts from different disciplines. Some have already been digitized and annotated. For example, Renn and colleagues in Florence have put on the Internet Galileo Galilei's *Notes on Motion* — four decades' worth of the Renaissance scholar's laboratory notes of studies into mechanical problems.

In cooperation with another team in Florence, working on a project called the Years of the Cupola, Renn is now building an electronic archive of over 14,000 administrative entries relating to the construction of the city's cathedral. One of the wonders of the Renaissance, this building features an enormous self-supporting cupola, or dome, 46 metres across at its base, that was completed in 1434.

The Archimedes project will examine how knowledge about mechanics evolved with time. A key question is how Aristotelian mechanics, based on 'intuitive physics', eventually lost its influence. For example, Aristotle equated motion with force, but when Galileo and other Renaissance scientists started to experiment and build up a formal theoretical basis for mechanics, they began to realize his error.

The cupola project is already yielding insights. It has, for example, uncovered pay slips for two 'observers' who were paid to watch architect Filippo Brunelleschi building a model of the cupola, to prove it could be safely constructed without expensive scaffolding. "This says something about a point in time at which mechanical knowledge was moving forward to new feats which had not been tested," says Renn — who expects such



serendipitous connections to become more frequent when automated search tools have more information to sift.

Other sources include engineering drawings for machines, analysed according to their elementary building blocks; images from ballistic treatises, analysed for their reliance on theoretical or practical knowledge (seeing, for example, what the authors recommend as the best angle of 'throw'); and dimensions of artefacts such as balances and standard weights, together with video records of their production and use.

Expert input

Annotating these diverse sources of information will require intellectual input from scores of historians, together with experts of other disciplines. Collaborators include physicists, archaeologists and even cognitive psychologists — who are working with children to help develop a better understanding of intuitive physics.

Aside from enlisting expert knowledge and developing computational tools, the main problems facing Renn and his colleagues centre on access to source materials. Copyright can be a thorny problem and archivists, who tend to see themselves as guardians of historical treasures, are sometimes obstructive.

Renn's digital revolution has also met resistance from some academic historians.

Fritz Krafft, head of the Institute for the History of Pharmacy at Marburg University in Germany, worries that historians working on translations by computer from a language with which they are not familiar could be led to false interpretations. John Heilbron, a historian of science at the University of Oxford, also remains unconvinced by a brute-force informatics approach. "You need new ideas, not just new tools," he says.

But many are enthusiastic. "This is the way scholarship is going," says Seamus Ross, director of the Humanities Advanced Technology and Information Institute at Glasgow University. "This approach makes humanities more science-like, because of the ease of repeatability and testability of arguments against other interpretations." Rudiger vom Bruch, a historian of science at the Humboldt University in Berlin, agrees: "Historians would be badly advised if they didn't use the possibilities offered by digitization."

Some historians may fear that Renn's approach will rob their discipline of its romance. But once Archimedes and its successors bear fruit, he argues that such fears will prove unfounded. "On the contrary, it will revitalize the potential of the humanities to reflect upon our culture," he says. ■

Alison Abbott is Nature's senior European correspondent.

Cuneiform Digital Library Initiative

♦ <http://www.cdli.ucla.edu/index.html>

Galileo's *Notes on Motion*

♦ http://www.mplwg-berlin.mpg.de/Galileo_Prototype/MAIN.HTM

Perseus Project

♦ <http://www.perseus.tufts.edu>

Changing patent laws could be a healthy move to combat resistance

Sir—A contradiction exists with the commercial development of chemotherapeutics such as antibiotics, anthelmintics and antiprotozoal agents, and the generation of resistance in organisms to these agents, for which I would like to suggest a solution.

Chemotherapeutic companies invest large amounts of money and effort in research, registration and marketing of new products. This investment is usually protected for 15 years after patent acceptance. However, by the time the product reaches the market the patent period has partially expired (even though in many cases companies are able to recoup investment through additional patent claims and similar patenting strategies).

In some cases the company needs to have optimized its return on investment by the time the patent expires and generic products appear on the market. This means that there is considerable commercial pressure on the company to achieve the greatest possible market penetration and sales volume to service the investment before this time.

As well, in the increasingly global free-market economy that we live in, chemotherapeutic companies are under an obligation to maximize profits for their shareholders, both in the short term and in the long term.

Chemotherapeutic companies have to promote their product in the most effective and efficient manner to ensure their own commercial survival and to maximize the return on their investment, while at the same time trying to husband the use of their products for the greater long-term public good.

This task is a paradoxical challenge for manufacturers, national registration agencies and governments. The use of these agents, as has been repeatedly demonstrated, results in the selection of existing resistant organisms within the population concerned, and/or creates favourable conditions for resistant mutants to thrive.

The cycle of short-term commercial pressure for optimal market return on investment and the resultant generation of resistance should be broken, if possible, in the long-term interests of human and animal health.

A radical and self-regulatory approach to break this cycle could be to link the continuation of a patent to the resistance status of the product. This approach

would encourage marketing strategies and patterns of use that are designed to preserve the effectiveness of the preparation and to minimize the development of resistance.

There are likely to be many and varied practical, theoretical and ideological objections to this proposal. Examples are that it is too difficult to change the patent laws and that the nature of the development of resistance is unpredictable.

Is it possible to address this paradox without a lateral approach, such as altering patent laws internationally and reducing the commercial pressure for short-term commercial return on investment?

Any new agents for the chemotherapeutic treatment of bacterial, parasitic and protozoal diseases need to be conserved, if possible, for future generations. To date, the evidence for longevity of the effectiveness of many existing drugs in medicine and animal health has not been good.

Terry Nicholls

11 Poole Place, Latham ACT 2615, Australia

We need both computer models and experiments

Sir—We were delighted by your News report “Computer modellers seek out ‘Ten Most Wanted’ proteins” (*Nature* **409**, 4; 2001), highlighting recent advances in protein structure prediction.

The report states, however, that computational structural biologists “believe that computer modelling may become a viable alternative to the current practice of using X-ray crystallography and nuclear magnetic resonance techniques to ‘solve’ the structure of a protein”.

The importance and value of structure prediction methods is to produce rough, approximate models for proteins with no experimental structure available. These models can often provide verifiable hypotheses about function, as your report accurately mentions. However, in the foreseeable future we expect predicted structures neither to approach the accuracy of experimentally determined structures, nor to replace them.

The report also states: “The big question, researchers say, is whether modelling will continue to improve fast enough to make an impact. There is a relatively short window of time for structure prediction to be useful before structural genomics ... generates all the results needed.”

We disagree strongly with this statement. Structural genomics aims not to solve the structures of all proteins, but

rather to obtain a set of representative structures such that all others can be modelled. Thus computational methods will play a critical role in translating the information on the relatively small fraction of proteins whose structures will be solved into accurate models for all proteins.

We furthermore expect that computational methods will produce useful low-resolution models for large numbers of sequence families many years before representative structures can be solved experimentally, and will play an all-important role in the design of novel proteins and therapeutics.

Daniel Fischer*, **David Baker†**, **John Moulton‡**

*Bioinformatics, Department of Computer Science, Ben Gurion University, Beer-Sheva 84015, Israel

†Department of Biochemistry and Howard Hughes Medical Institute, Box 357350, University of Washington, Seattle, Washington 98195, USA

‡Centre for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, Maryland 20850, USA

Why Pauling didn't solve the structure of DNA

Sir—Linus Pauling, who was celebrated in Gautam R. Desiraju's Millennium Essay (*Nature* **408**, 407; 2000), was born on 28 February 1901. As the hundredth anniversary of his birth approaches, I would like to describe Pauling's comments on the discovery of the structure of DNA.

Shortly before Christmas 1988, Linus was the keynote speaker at the UCLA–Sloan winter school on molecular evolution. He struck up an immediate rapport with my wife, Laura, showing her pictures of his late wife and of his son as a toddler. As we were leaving his lecture to take him back to the faculty guest house, he suddenly asked my wife and me in his uniquely direct way if we ever wondered “why he hadn't solved the structure of DNA”.

We were totally surprised by this, but at the same time were curious to know the answer and asked him why. He said that one day his wife had asked him that question. It had made him think and he replied something to the effect of “I don't know, I guess that I always thought that the DNA structure was mine to solve, and therefore I didn't pursue it aggressively enough”.

Perhaps this isn't the only reason, but on that December night it was Linus's reason. His presence is missed.

Jim Lake

Molecular Biology Institute and MCD Biology, UCLA, Los Angeles, California 90095, USA

Food for thought ... and action

Technology that could transform food production stands at a crossroads.

Pandora's Picnic Basket: The Potential and Hazards of Genetically Modified Foods

by Alan McHughen

Oxford University Press: 2000. 277 pp.

£16.99, \$25

Food's Frontier: The Next Green Revolution

by Richard Manning

North Point Press: 2000. 225 pp. \$24

Dick Taverne

The debate in Britain about genetically modified (GM) food, which has attracted international attention, has not been a rational debate about the evidence. The public has been scared by headlines about "Frankenstein foods". It increasingly distrusts scientists, particularly since the BSE débâcle, and prefers to rely on information from pressure groups such as Greenpeace and Friends of the Earth. These in turn seem more interested in publicity to boost membership than in examining the evidence on its merits. Meanwhile, the press has largely ignored attempts at objective analysis, such as the reports from the Royal Society, select committees of both Houses of Parliament and the admirable and comprehensive study by the Nuffield Council on Bio-Ethics.

Yet the evidence clearly suggests that scares about the safety of GM foods are unfounded. A conference on this subject was held in Edinburgh a year ago by the Organisation for Economic Co-operation and Development. It was attended by some 400 experts from all over the world, as well as representatives of the pressure groups. At the end, the chairman, John Krebs, asked whether anyone could cite any evidence to show that GM food or crops were a threat to health. There was a prolonged and telling silence.

Alan McHughen, the Canadian author of *Pandora's Picnic Basket*, was one of the speakers at the conference. His book is essentially a layperson's guide to the subject and would make an excellent textbook for schools. It is cogently argued, clearly written and intelligible to a non-scientist such as myself.

In one respect he misleads: he claims to chart a middle course between the extreme claims of the proponents and opponents of GM technology and to provide a balance between the potential risks and benefits. In fact, he is hard pushed to name the risks. There could be a risk to health if an allergenic protein were inserted into a novel food, but McHughen is confident that it would be eliminated long before it reached his kitchen. About the only potentially dangerous example he cites is the insertion of a brazil nut gene



Last best hope? GM rice with a vitamin A gene inserted could enhance nutrition in poorer countries.

into a soya bean. Actually, no such bean was ever marketed. It was withdrawn 10 years before anti-GM activists were aware of the danger. As for risks to the environment, McHughen suggests that we should treat some GM crops now under development with special care. When a pesticide-resistant crop 'escapes' into the wild, it has no competitive advantage, as wildlife is not sprayed with pesticides. But genetically engineered stress tolerance might create problems — a GM plant able to tolerate freezing conditions could have a considerable competitive advantage over its neighbours in the wild.

McHughen focuses the light of reason mainly on current myths, with devastating effect: that a transgenic plant, or gene

transfer, is 'unnatural'; that conventionally farmed food, or even organic food, is safer; that transgenic crops are more dangerous to biodiversity or are likely to create 'superweeds'; that labelling, especially of so-called 'GM-free' products, gives the consumer a meaningful choice; and that, in testing food, we should be concerned with the process rather than the product.

Surprisingly, one subject he underplays is the potential of transgenic crops to fight hunger and disease in the developing world. There is only a cursory mention of the enormous benefits to children in the developing world that could flow from the insertion of a vitamin A gene into rice. This is the kind of GM development that it seems wholly

book reviews

immoral to oppose and which clearly discomforts opponents of the new technology.

Food in the developing world, and what happens now that the green revolution is running out of steam, is the topic of Richard Manning's book. His is a very different approach, personal and anecdotal where McHughen is rational and analytical. It is an account of his experiences in nine countries in Africa, Latin America and Asia and the lessons to be learned.

Manning confesses that he came to the issue of transgenic crops almost as an afterthought. He began the book, he writes, largely ignorant of it, inclined to side with those who counselled caution before rushing into this brave new world. But he found it was almost ubiquitous in the agriculture of the developing world, and became convinced of its potential benefits. He finds that, at a deeper level, "it is helping to rewrite our understanding of life".

But most of his book has a different message: how much can be done and needs to be done at a less-sophisticated technical level to feed the developing world. The application of GM technology is one thing in China, where scientific expertise is widespread and growing, and another in Uganda, where expertise is limited to a handful of scientists.

Manning relates how, in many countries, such as Zimbabwe and Peru, the old lessons of good practice have been forgotten; and elsewhere simple improvements were never learned. Yet the revolution that must take hold where yields have stagnated must build on local practice and local communities. "The green revolution at its most fundamental level treated all the world the same. The lessons being learned in agriculture now are local."

Curiously, the conclusion I draw from these two very different books is that they complement each other. Transgenic crops are likely to play a vital part in feeding the extra two billion people expected to be born in the next quarter-century, but only if it is applied in a way that enhances local communities, fits in with local practices and cultures and recognizes the need for complexity and diversity. There is the danger that GM technology could lead to a greater degree of monoculture, as some of the non-governmental organizations (NGOs) fear. But it could, with their cooperation, mean that the next green revolution avoids the uniform approach of the last one and the damage this caused to many small farmers. In time, no doubt, the NGOs will reconcile themselves to the inevitability of the new technology. Once they do so, their dedication to improving the life of the poor can help to ensure that transgenic crops realize their huge potential benefits. ■

Dick Taverne is a Liberal Democratic member of the UK House of Lords, and has a special interest in the public understanding of science.

Getting the point across

Acupuncture: Efficacy, Safety and Practice

edited by the British Medical Association
Harwood: 2000. 100 pp. £12.99, \$24 (pbk)

John Galloway

It is a good thing to be reminded that, although medicine tends to be regarded as a learned profession, it is perhaps better thought of as a collection of craft skills, and an eclectic one at that. Anything goes; well, almost anything. Whatever you learn about acupuncture from this book by the British Medical Association (BMA), that is the take-home message. Not that there is necessarily anything wrong with an eclectic approach; indeed, there is a lot to be said in its favour. People want different things from medicine and ought to get them. There is, of course, the question of whether the medicine on

offer actually works or, at least, does you more good than harm.

What the BMA has done is put together a short introduction to the subject. What exactly is acupuncture? Does it work? Is it safe? Where can you get it done? What is the attitude of the UK National Health Service to acupuncture? And so on.

The first lesson to be learnt, and it is a good one, is that acupuncture comes in two guises. First, there is what might be called the traditional guise, as an integral part of the very old system of Chinese medicine, which some archaeologists believe has its roots in the Stone Age. Then there is a modern, 'scientific' guise, which finds its theoretical rationalization in the "gate theory" of pain devised by Ronald Melzack and Patrick Wall in 1965.

It is worth examining the two approaches because, however far apart they are philosophically, the actual treatment looks remarkably similar, at least to the untutored eye. Chinese medicine is based on the need to achieve balance between opposing



Needed: acupuncture, long practised in China, targets specific points on the body's energy flow paths.

CORBIS

LISA M. MCGEADY/CORBIS

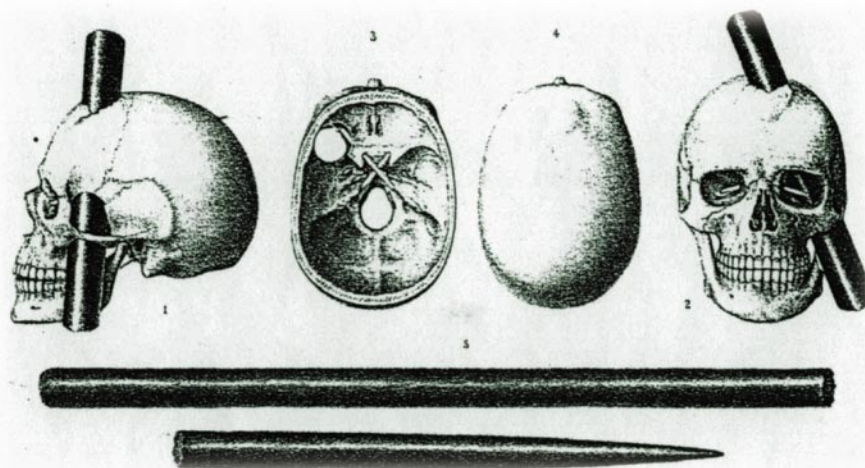
principles, or 'forces': cold and dark against warmth and light; positive against negative, for instance. There is also a second strand involving 'elements': earth, fire, water, wood, metal — reminiscent of the humours of Ancient Greek medicine. Chinese medicine is comprehensively holistic, treating the body, mind and spirit as indivisible and a person as inextricable from his or her environment. All in all, it is a subtle and fluid system, and hardly seems susceptible to analysis in terms of Western, deterministic, reductionist science.

The gate theory, at its simplest, suggests that the stimulation of some nerve fibres that transmit pain impulses is responsible for blocking pain impulses in other nerve fibres. You can see the attraction of the theory as an explanation of how acute pain might be relieved by stimulating nerves with needles. Indeed, the Royal Society, providing evidence for the House of Lords Inquiry into Complementary and Alternative Medicine, reported that meta-analysis of clinical trials of acupuncture for treating pain had shown a beneficial effect — although they added that they suspected publication bias. Of course, acupuncture is not traditionally used exclusively for treating pain. Nevertheless, the gate theory of pain seems to have given acupuncture some scientific respectability and to have led to a variant of the traditional use in the form of 'scientific acupuncture' for treating acute pain.

Does acupuncture actually work? According to the book, it is surprisingly well researched, at least in the number of trials of its clinical effectiveness, if not always their reliability. Rigorous systematic reviews are the backbone of the evidence-based medicine movement. The book describes trials of acupuncture for nine conditions, including back, neck and dental pain, recurrent headaches, weight loss, nausea and vomiting, recovery from stroke, stopping smoking and osteoarthritis. This is a pretty mixed bag and it seems fair to say that success is also mixed. There is evidence that acupuncture works to some extent in controlling post-operative sickness and for easing neck and dental pain — but remember the Royal Society's caveat of publication bias. However, it is virtually useless for the other conditions assessed.

Yet acupuncture is popular in Britain, and the number of its practitioners is growing rapidly. In a BMA survey of its members, 80% of doctors in general practice said that acupuncture should be available in the National Health Service, half of them saying that this was because of its "proved effectiveness". The interesting insight, however, was the large number of doctors who felt that they should be the ones to give the treatment, and their strong antipathy to allowing traditional Chinese practitioners to be involved.

And there was no sense that doctors feel the need to embrace the whole of the Chinese



Run through: sketches showing the fateful consequences of Gage's momentary lapse of attention.

tradition of medicine, in which acupuncture plays a relatively minor part, or that they consider the philosophy behind the treatment as important. Acupuncture is clearly seen as a technical skill that could be learned in isolation and added to the stock of orthodox medicine, with its completely different philosophical foundations. Pragmatic, yes. Business-like, certainly. But the hallmark of a learned profession? I am not so sure. Perhaps what the book lacks is a point of view. In trying to be balanced it gives an impression of lacking in firmness of purpose. Come on, BMA! On the grounds of philosophical plausibility and the available evidence, yes or no? ■

John Galloway is at the Eastman Dental Hospital, 256 Gray's Inn Road, London WC1X 8LD, UK.

New in paperback

What is Life?

by Lynn Margulis & Dorion Sagan
University of California Press, \$24.95, £16.50

The Dilemmas of an Upright Man: Max Planck and the Fortunes of German Science

by J. L. Heilbron
Harvard University Press, \$16.95, £11.50

Out of the Crater: Chronicles of a Volcanologist

by Richard V. Fisher
Princeton University Press, \$14.95, £9.95

The Nazi War on Cancer

by Robert N. Proctor
Princeton University Press, \$16.95, £10.50

Managing Sustainable Development

by Michael Carley & Ian Christie
Earthscan, £15.95

Legacy of a flying tamping iron

An Odd Kind of Fame: Stories of Phineas Gage

by Malcolm Macmillan
MIT Press: 2000. 562 pp. \$39.95, £27.50

Ian Glynn

"This book tells me more about penguins than I want to know," said the little girl when asked whether she liked her present. At first sight, a book of over 500 pages about Phineas Gage and the event that made him famous might seem in danger of similar criticism. But Malcolm Macmillan's enthusiasm is as infectious as his knowledge of detail is prodigious. And if we are sometimes told more than we want to know, what we are told is often fascinating, and answers questions that — a century and a half after the event — we would not expect to be answerable. But, first, who was Gage and what was the event?

Phineas P. Gage was the 25-year-old popular and shrewd foreman of a gang helping to build the Rutland and Burlington Rail Road in Vermont. A momentary lapse of attention, while he was blasting rock on a sunny September afternoon in 1848, exploded the charge prematurely, converting a tamping iron into a missile that entered his left cheek, passed upwards through his skull and landed "several rods" away.

Astonishingly, Gage was not killed, or even long unconscious, and after less than three months under the care of Dr John Harlow he had recovered sufficiently to go home. But although, two years after the accident, Henry Bigelow, the professor of surgery at Harvard University's medical school, described Gage as "quite recovered in his faculties of body and mind with the loss only of the sight of the injured eye", his

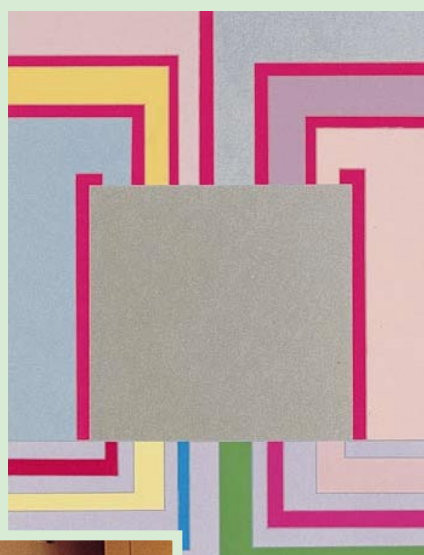
personality had changed drastically. His contractors, who had regarded him as their most capable foreman, were unwilling to re-employ him; and Harlow, writing much later, describes him as "fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom) ... impatient of restraint or advice ... at times pertinaciously obstinate yet capricious and vacillating, devising many plans of future operations, which are no sooner arranged than they are abandoned in turn for others more feasible". The balance between Gage's intellectual faculties and animal propensities, Harlow felt, had been destroyed.

For some years Gage wandered round the larger New England towns doing a variety of jobs, and even spending some time at Barnum's Museum in New York, complete with his tamping iron. He then spent nearly eight years in Chile looking after horses, and driving a six-horse coach, before returning to his mother and sister in San Francisco, where he died after a series of fits in 1861.

For 150 years, Gage's accident and its consequences have been discussed in newspaper articles, learned journals, lectures, neurological conferences, medical textbooks and, more recently, popular television programmes; it has also spawned fiction. All this material is grist to Macmillan's mill. What emerges is a remarkable series of essays — detective stories without criminals — whose scope is far wider than the title of the book suggests, and whose interest switches between the personal, the historical, the medical and the neuropsychological, reflecting the varied enthusiasms of the author. What sort of family did Gage come from? Was he a subcontractor, or just an employee? And what kind of men made up the gangs involved in blasting? What was Dr Harlow like and what was the role of Edward Williams, the young doctor whom Gage saw half an hour after the accident and greeted with the words, "Doctor, here is business enough for you"?

How could a man survive after an iron bar, three feet seven inches long and weighing thirteen and a quarter pounds, had been driven through his head? Did Gage recover in spite of or because of Harlow's treatment? How could Bigelow, at Harvard, have been so wrong in his assessment? How could so many of those who subsequently discussed Gage's case have been so confused? Was it the result of failure to read the not very accessible contemporary accounts, or of bias towards particular views on the

Science in culture



Artists' impressions of art Interpretations of medical illustrations.

Alison Abbott

Towards the end of his long and extraordinarily productive career, Frank H. Netter was referred to by *The New York Times* as the Michelangelo of medicine. This may confuse skill with genius, but Netter's medical illustrations, which are

immortalized in the 13-volume *Netter Collection of Medical Illustrations* published by Ciba (now Novartis), show a talent and humanity that have been appreciated by medical students for decades. Netter's aim was to illustrate not just the anatomical aspects of a disorder, but also its emotional aspects, through appropriate facial expressions and body language.

Twenty-five years ago, Ciba agreed for nine of the volumes to be translated and published for the German-speaking market. Permission for publication of the last four volumes, whose illustrations publisher Thieme wanted to modify for scientific accuracy, has been negotiated more recently and is under way.

localization of function in the brain? Did biases in favour of or against phrenology play a part? What does Gage's case tell us about localization of function in the cerebral cortex? Did Gage's case play an important or even a significant part in the development of psychosurgery? Was Freud influenced by it?

The book's success lies in the combination of Macmillan's skill as a writer, his familiarity with the labyrinthine development of nineteenth-century ideas about the

Farbatlanten der Medizin

The Ciba Collection of Medical Illustrations

Band 1:

Herz

Konzept und Illustrationen:

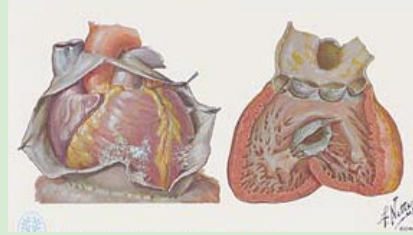
Frank H. Netter

Redigiert von Fredrick F. Yonkman

Übersetzt von Krista Schmidt

Herausgegeben und überarbeitet von Martin Stausch

271 farbige Tafeln
3. überarbeitete und erweiterte Auflage



Netter's cover for his volume on the heart (above) and the response to this by Peter Halley (left).

In celebration of the fact that all 13 volumes will be published in German, Thieme invited 10 renowned contemporary artists to interpret Netter's work. Each of the artists received one of the volumes — 'nervous system', 'kidneys', 'genital organs', and so on — to respond to. The artists, who include Rosemarie Trockel, Lawrence Weiner, Peter Halley and Andres Serrano, appear to have been selected on the basis of fame and variety of style, rather than their use of the same medium as Netter (few are painters) or previous involvement with anatomical themes. Their interpretations are presented in the newly published tenth volume of the German version of the Netter collection, along with essays and commentaries. Samples of the works are touring galleries in Germany, Austria and Switzerland this year.

Alison Abbott is Nature's senior European correspondent.

Exhibitions planned so far:

8–10 February, Raab Gallery, Berlin;

29–31 March, Galerie Hohenlohe & Kalb, Vienna;

24–28 April, Galerie Schüppenhauer, Cologne.

♦ <http://www.thieme.de/netter-art-collection>

brain and his passion for collecting and presenting evidence, whether historical or scientific. Of course, not all of his arguments are equally convincing, but only rarely is one seriously worried about his conclusions. And if the accounts of Alexander Bain's early life and of Harlow's later life lead to thoughts of penguins, both were admirable men and their biographies occupy only a very small fraction of the book.

Ian Glynn is at Trinity College, Cambridge CB2 1TQ, UK.

Back to the future from 1888

One writer's vision of a technological utopia has stood the test of time.

Howard P. Segal

Over the centuries, the designated route to utopia has taken many forms. These include various economic, religious and sexual beliefs and practices. Scientific and technological advances have been no less popular, especially as the Scientific Revolution and English Industrial Revolution made progress a fact of daily life rather than a mere fantasy. Likewise, it is no accident that the questioning of 'progress' in the twentieth century is epitomized in such anti-utopian works as Eugene Zamiatin's *We* (1920), Aldous Huxley's *Brave New World* (1932) and George Orwell's *1984* (1949). The ability of modern science and technology to change the world is not at issue in these writings. Instead, it is the fear that they will be all too successful.

Edward Bellamy's *Looking Backward: 2000–1887* (1888) is, by contrast, an avowedly utopian work, looking positively towards the future. It is a product of North America's, and Victorian Britain's, peak of faith in science and technology as panaceas. Among the most popular utopian novels ever published, *Looking Backward* remains available in several editions. Every day, new readers are transported along with Julian West — a wealthy young Bostonian who somehow sleeps for 113 years — from the poverty, disease, violence and corruption of late-nineteenth-century industrial America to a nation of well-planned cities, full employment, material abundance, good health and social harmony. In the process, they join Julian in his quest to understand how utopia emerged without bloodshed or revolution. As hundreds of other utopian writings in the late nineteenth and early twentieth centuries have faded into obscurity, *Looking Backward*'s enduring popularity is ever more striking.

One neglected explanation for this popularity is that *Looking Backward* is as much about technology and its use and abuse as it is about cooperation and semi-socialism — the focus of most analyses. Technology is at once a principal cause of and solution to the major problems Bellamy addresses: inefficient production; inequality of opportunity and of income; immorality; monopoly and exploitation; and urban blight. Technology's purposeful, positive use — improved factories and offices, new highways and electric

lighting systems, innovative pneumatic tubes, electronic broadcasts and credit cards — is critical to the country's predicted transformation from living hell to heaven on Earth.

Indeed, Bellamy's envisioned twenty-first-century United States is a technological utopia: a society dependent upon machines and openly modelled on them. Bellamy grasped technology's potential to create a reformed social and civic order out of what, to many Americans, was material and moral chaos. He acknowledged the value of earlier utopian communities such as Massachusetts' famous Brook Farm, but argued that the modern United States, with its urbanized and diversified population, required different models for improvement. Blueprints for the future should now take the form of readable books, not small escapist settlements. And Bellamy (1850–98), a professional journalist and accomplished fiction writer, knew how to captivate readers.

Far from being a rigid technocrat demanding complete conformity, Bellamy was exceptionally sensitive to technology's potential for satisfying citizens' needs and desires. One overlooked example is *Looking Backward*'s attention to technology-based centralization and decentralization alike. On the one hand, we find centralization in the overall administration of society, especially the 'industrial army' commanded by the president; in production, distribution, and consumption facilities such as department stores and warehouses; in domestic enterprises and in huge urban apartment complexes. On the other hand, we see decentralization in widely separated towns and villages; in electronic broadcasting systems sending music from music-halls and sermons from churches to home telephones; and in dining facilities within individual apartments. There is no conflict here between centralization and decentralization, but rather balance and variety.

Ironically, Bellamy predicted only a few specific technological advances: those pneumatic tubes to deliver goods, electronic broadcasts and credit cards instead of cash. Hence the critique of *Looking Backward* by some recent professional forecasters as a prophetic failure. Yet several social critics indict *Looking Backward* for being too accu-



A man of the future: Bellamy's utopian vision considered the use and abuse of technology.

rate. They condemn its "scientific management" as a precursor of totalitarian communism and fascism and so connect the book to anti-utopian visions such as *1984*. But this distorts Bellamy's basic optimism, his far from radical vision of middle-class respectability, and the choices of occupation, residence and leisure he offers citizens. Indeed, few citizens work after 45; marriage is for love, never wealth, because everyone lives comfortably; and crime barely exists because poverty no longer exists. The Nationalist movement growing out of the book's extraordinary popularity wished to reform the United States' political system, not to overthrow it.

What does link Bellamy to the twentieth-century anti-utopians such as Zamiatin, Huxley and Orwell is that each wrote out of moral conviction. So, too, have all major utopian writers from Thomas More to Buckminster Fuller. By contrast, the most popular contemporary futurists write and routinely rewrite for the market-place. These self-proclaimed visionaries seduce a culture that, beneath surface sophistication, seeks simple reassuring answers to complex unsettling questions.

So if Bellamy anticipated our high-tech advances only in credit cards, he is nevertheless still timely in his concern for using technologies efficiently, for distributing the wealth derived from technological advances fairly, for offering varieties of work and living arrangements, and, not least, for writing passionately about the future. One need not be a professional prophet to appreciate this. ■

Howard P. Segal is at the University of Maine, 5774 Stevens Hall, Orono, Maine 04469-5774, USA.

CORBIS/ARY EVANS

A never-ending story

Peter D. Moore

Nothing succeeds, it seems, like succession. A concept that emerged from the sand dunes of Lake Michigan in 1889 has grown beyond recognition and has expanded into almost every area of ecological thinking. Succession's existence is now generally accepted, despite continuing disputes about its underlying mechanisms. Rather like the physics of light, we need to accept that there are many ways of looking at this universal but enigmatic process. Ignorance does not prevent us from applying the concept of succession to the manipulation and management of ecosystems, either for the production of consumables or for wildlife conservation.

Henry Chandler Cowles first developed the idea of succession in his studies around the shores of Lake Michigan, where the gradual shrinking of the lake since the last glaciation had progressively exposed new terrestrial habitats. The developing dune ridges represented a series of stages in the colonization of the new land. Vegetation changed from initial specialist grasses to poplars (cottonwoods), then to pines and finally the oak forest that forms the region's 'climax' vegetation. Cowles, very reasonably, took the spatial arrangement of these communities of plants to reflect their progressive develop-

ment through time. And herein lies one of the great problems in the study of succession, for the timescale of this type of vegetation development (centuries, perhaps running into millennia) is far beyond the lifespan of humans, so the process must be followed by proxy rather than by direct observation.

An important feature of succession is its predictability. Ecosystems pass through a series of stages and eventually stabilize at a predetermined end-point, the climax. This point is essentially controlled by climate, although other factors such as soil conditions may lead to local variation. But the stages may simply be artefacts of the time dimension, mere snapshots in a continuous process. Similarly, the impression of a series of communities may be a mental imposition determined by our temporal limitations: species of plants, animals and microbes may actually arrive and depart in isolation rather than in concert.

The apparently deterministic nature of succession could be due to our own need to seek pattern. Indeed, studies of wetland successions, which leave fossil records of their courses in stratified lake and peat sediments, show that various pathways of development are possible. Stochastic elements, including the availability of certain immigrant organisms, may affect the line that is taken. The existence of an end-point, an ultimate stable equilibrium that is self-sustaining, is increasingly unacceptable as we come to understand the pattern and pace of environmental changes, from catastrophes to climatic shifts. On a scale of centuries or millennia, the Earth will have changed beneath an ecosystem's roots and a new set of conditions will place fresh demands on the turbulent vegetation.

A kind of anthropomorphism has also attended the development of the concept of succession, simultaneously stimulating and blighting it. Frederic E. Clements, at the close of the nineteenth century, and Sir Arthur Tansley, at the beginning of the twentieth, encouraged an organismal view of ecosystem development. They saw succession as a type of embryology leading ultimately to the fully developed ecosystem in its final maturity.

This extreme form of determinism has failed the test of time. The species of plants and animals that are involved in successions arrive when conditions are appropriate and when their populations are geographically available. They depart when conditions are less suitable or when competition for resources proves too severe. Species, as Henry Gleason pointed out in the 1920s, act in an individualistic manner; communities are simply the assemblages of the moment. Palaeoecological studies of plant and animal movements during the past 14,000 years

Succession

"Species act in an individualistic manner; communities are simply the assemblages of the moment."

have supported the idea that it is individual species, rather than communities, that move in response to climate change.

In 1969, Eugene P. Odum set out an energetic model for succession by concentrating on the general features of the process. Energy balance in the ecosystem progressively changes, with ecosystem respiration lagging behind production. When the two eventually coincide, equilibrium — climax — is attained. Biomass is generally greatest at this equilibrium stage, nutrient imports to the ecosystem are equalled by exports, and species richness and general complexity are at their peaks. The model has many attractive and useful features, although the final state of equilibrium must still be regarded as a dubious ideal.

What mechanisms underlie the observed sequence of species? Each species may modify the environment in some way, rendering it more suitable for the invasion of others — as in the provision of soil organic matter, soil stability, nitrogen fixation, shade, and so on — processes that have been called facilitation. Or perhaps competitive factors predominate, early colonists exploiting a situation of low competition, only to be displaced as more aggressive but slower invaders arrive. It is likely that both processes are present and interact together.

As in so many areas of science, a lack of full understanding of a process does not necessarily interfere with its exploitation. As we manage ecosystems, we often set succession into reverse gear. We create open water in wetlands, open glades in forests, extract timber, plough land and take harvests from annual plants such as wheat. All these activities are manipulations of the process of succession. Whether seeking enhanced biodiversity or agricultural productivity, we often need to keep succession in check. ■

Peter D. Moore is in the Division of Life Sciences, King's College, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NN, UK.

FURTHER READING

Krebs, C. J. *Ecology: The Experimental Analysis of Distribution and Abundance* 5th edn (Addison Wesley Longman, San Francisco, 2001).

Luken, J. O. *Directing Ecological Succession* (Chapman & Hall, London, 1990).

Odum, E. P. *Ecology: A Bridge Between Science and Society* (Sinauer, Sunderland, MA, 1997).

Shugart, H. H. *Terrestrial Ecosystems in Changing Environments* (Cambridge Univ. Press, 1998).

Father of a successful idea: Henry Cowles.



Fusion needs more than SNAREs

Wolfhard Almers

Force biological membranes close enough together and they will fuse. SNARE proteins are well suited to force proximity. But biochemical studies of yeast show that proximity is not the only requirement.

The cells in our bodies are packed full of membrane-bounded compartments. These compartments exchange material by shedding small, cargo-carrying packets called vesicles, which then fuse their membrane with that of other compartments or with the plasma membrane. Without fusion, all such 'membrane traffic' would grind to a halt. Nerve cells would stop releasing neurotransmitter, and cellular compartments would shed vesicles until the compartments disappeared.

It seemed all but settled that membrane fusion was mediated by the so-called SNAREs — proteins that inhabit all fusion-competent membranes of a cell. But a paper on page 581 of this issue¹ threatens to overturn that view. Peters and colleagues report a situation in which SNAREs do not cause fusion directly, but instead enable an ion-permeable membrane protein to form a bridge between membranes. This then allows fusion to proceed.

SNAREs have been studied extensively in neurons and other secretory cells, where these proteins help secretory vesicles to fuse with the plasma membrane. When two membranes fuse², SNAREs on the vesicle membrane (v-SNAREs) reach out to grab hold of t-SNAREs on the target membrane. The SNAREs then intertwine in a tight 'trans' complex, which is thought to cause fusion by pulling the two membranes to within 4 nanometres of each other³. After fusion, v-SNAREs and t-SNAREs unite their anchors in the fused membrane, so forming a 'cis' complex. Such complexes are so stable that energy is needed to disassemble them for re-use. The energy for this comes from the hydrolysis of ATP (adenosine triphosphate) by a protein called NSF.

Fusion cannot occur without SNAREs — it is blocked when SNAREs are removed genetically or destroyed by neurotoxins. Moreover, purified SNAREs alone can cause fusion when reconstituted into liposomes⁴, which are artificial membrane-bounded spheres. Most researchers in this field hold that several trans-SNARE complexes form a ring lying flat between the fusing membranes, and that fusion results from the forced proximity of membranes within the ring and from the stress exerted by the SNAREs. Lipids inside the ring are thought to leave their own bilayers and enter into temporary structures with lipids from the other bilayer, leading to fusion^{4,5}. But exactly how trans complexes cause fusion is unclear, and the temporary structures are still hypothetical.

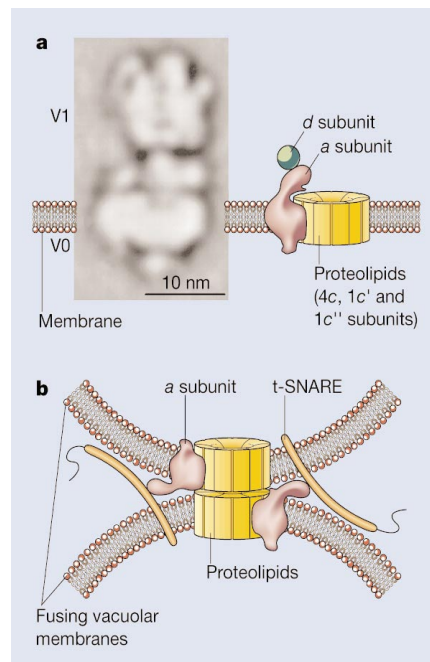


Figure 1 Subunits of the vacuolar proton pump. **a**, Left, electron micrograph of the whole mammalian enzyme (reproduced from ref. 11), showing its V0 (embedded in the membrane) and V1 sectors. Right, a diagram of V0. Its ring of proteolipids consists of four *c* subunits, encoded in yeast by the *VMA3* gene, and one each of subunits *c'* and *c''*, encoded by yeast *VMA11* and *VMA16*, respectively. The *a* subunit (encoded by *VPH1*) embraces the ring. The position of the *d* subunit (encoded by *VMA6*) is unknown. During proton pumping, the proteolipid ring and the *a* subunit rotate against each other, with proton transport occurring at the interface. The proton-transport pathways are shown schematically as vertical grooves. The central cavity probably does not conduct ions. **b**, Possible structure of the dimer of proteolipid rings (V0) that may form during vacuole fusion¹. Two rings of proteolipid, two *a* subunits and two target-membrane (t) SNARE proteins are shown. The calcium-activated protein calmodulin and the *d* subunit of V0 (not shown) are also associated with the dimer.

Studies of yeast vacuoles suggest a different picture. Vacuoles are large, fluid-filled cavities in cells, and, in yeast, they digest proteins. Vacuole fusion requires trans-SNARE complexes, but these apparently act early on and can disassemble before fusion occurs⁶, provided that the vacuoles have signalled their intent to fuse by releasing calcium ions⁷. So vacuole fusion was attributed to an unknown macromolecule that is activated by calcium and calmodulin (a calcium-activated enzyme) after the trans-SNARE complexes had formed.

In a biochemical *tour de force*, Peters *et al.*¹ now provide evidence that this macromolecule is the free V0 part of the vacuolar proton pump, which is found in most intracellular membranes and vesicles. The pump contains two multisubunit sectors — V1, which hydrolyses ATP, and V0, through which protons are transported (Fig. 1a). V0 consists of a complex of six so-called proteolipids, together with associated subunits. V1 and V0 can exist separately, but must combine to pump protons. Without V1, V0 becomes impermeable to protons. But the proteolipid ring of V0 can become permeable to the neurotransmitter acetylcholine when the proteolipid ring is reconstituted into liposomes and calcium ions are added⁸. Calcium also causes some ion channels to open water-filled holes that can pass acetylcholine. And calcium causes a

water-filled hole to open between the inside of a vesicle and the outside of a cell when acetylcholine is released from a neuron. The hole (called a fusion pore) is an early stage in fusion.

In yeast vacuoles, V0 apparently has two roles that can be separated experimentally. When bound to V1, V0 acidifies the inside of the vacuole. By contrast, free V0 helps membrane fusion to occur¹. Peters *et al.* show that two vacuoles fuse by forming a head-to-head dimer of V0 sectors (Fig. 1b), presumably at the point where the SNAREs first forced the membranes close together. The dimerization is triggered by calmodulin and raised concentrations of calcium ions. Calcium-bound calmodulin also causes the proteolipid complex of V0 to become permeable to choline, a close relative of acetylcholine.

Peters *et al.*¹ propose that dimerization allows two proteolipid hexamers to form a closed-off channel from one vacuole into the other (Fig. 1b). This structure suggests a model for fusion (Fig. 2, overleaf)^{1,9}. As well as causing dimerization of V0, calcium-bound calmodulin is envisaged to cause the proteolipids to separate within the plane of the membrane, so that lipids invade the spaces between them and the aqueous channel in the centre opens. The separating proteolipids expose between them surfaces that have a hydrophobic portion sandwiched

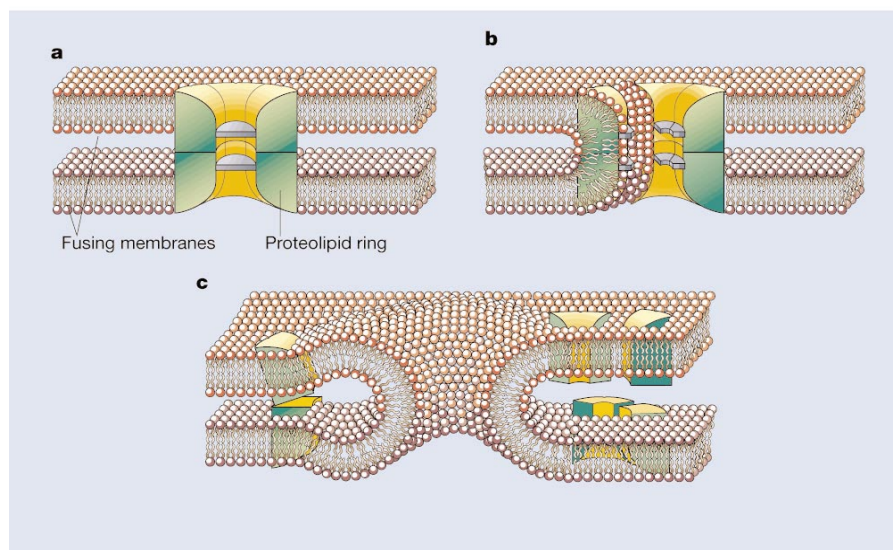


Figure 2 A model for the fusion of yeast vacuoles, as proposed by Peters *et al.*¹. First, the two fusing membranes are pinned together by SNARE proteins (not shown). **a**, A dimer of proteolipid rings (V0) forms. Yellow indicates hydrophilic and green, hydrophobic surfaces. Lines mark the borders between individual proteolipid monomers. The central cavity of each ring is closed off. **b**, The ring opens as the proteolipid subunits separate laterally. An aqueous channel — the fusion pore — forms in the centre. At the same time, lipids invade the spaces between the subunits. **c**, The fusion pore expands, ultimately causing the proteolipids to separate vertically. One problem with this model, however, is that the faces of the cavity in **a** end up facing lipid. This would be energetically unfavourable if the cavity were hydrophilic, as expected.

between two hydrophilic stripes. Such surfaces mimic lipid bilayers and allow lipids to invade the proteolipid ring without leaving their bilayer configuration. This model is attractive because it explains fusion without invoking temporary lipid intermediates⁹.

Do the proteolipid complexes dimerize directly, or indirectly through their associated subunits (Fig. 1a)? If the former, how are these subunits accommodated in the complex? Do proteolipids indeed separate laterally to form an aqueous channel? And what is the significance of the lone t-SNARE that Peters *et al.* found in the dimer of proteolipid complexes? Does it signal the break-up of the *trans*-SNARE complex before fusion, and does this break-up require NSF? Wickner, Mayer and colleagues^{6,7} have developed a powerful system that seems well suited for addressing such questions.

From a physiological viewpoint, why do vacuoles wait for the formation of a V0 dimer when SNAREs alone can cause fusion of liposomes? A similar question is why protein-based water channels exist in membranes, when lipid bilayers alone pass water. The answer to this question — increased speed and tighter control — may also apply to fusion. We don't yet know whether fusion events other than vacuole fusion rely on the proteolipid rings. But increased speed would be particularly useful in the case of neurotransmitter release — neurotransmitter-containing synaptic vesicles fuse with the plasma membrane less than 1 millisecond after neurons are stimulated. By contrast, one round of SNARE-mediated fusion of recon-

stituted liposomes takes 30 to 40 minutes⁴.

The presence of a ring of proteolipids at the fusion site may also make it easier for a vesicle to close its fusion pore soon after releasing its contents, preventing the membrane constituents from mixing with the target membrane. Closing fusion pores have been observed during most kinds of secre-

tion. After closing its fusion pore, the vesicle would be free to disconnect from the plasma membrane and prepare for re-use elsewhere. A proteolipid-based ring may also allow the flux of contents through fusion pores to be controlled, as proposed for hippocampal neurons¹⁰. And it may add an element of safety: fusion intermediates lacking sufficient protein lead to hemifusion⁵, an unproductive association between membranes that fails to allow the release of cargo.

It will be interesting to see further work on the role of proteolipids in fusion, and in particular whether proteolipids are also involved in the fusion of secretory vesicles with the plasma membrane. Many years ago, Israel and colleagues⁸ proposed that neurotransmitter is released not from vesicles but through a V0 proteolipid complex, which they called the mediophore. It would be poetic if they were now proven partially correct.

Wolfhard Almers is at the Vollum Institute, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97201-3098, USA.

e-mail: almersw@ohsu.edu

1. Peters, C. *et al.* *Nature* **409**, 581–588 (2001).
2. Brunger, A. T. *Curr. Opin. Neurobiol.* **10**, 293–302 (2000).
3. Jahn, R. & Südhof, T. C. *Annu. Rev. Biochem.* **68**, 863–911 (1999).
4. Nickell, W. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 12571–12576 (1999).
5. Chernomordik, L. V., Frolov, V. A., Leikina, E., Bronk, P. & Zimmerberg, J. *J. Cell Biol.* **140**, 1369–1382 (1998).
6. Ungermann, C., Sato, K. & Wickner, W. *Nature* **396**, 543–548 (1998).
7. Peters, C. & Mayer, A. *Nature* **396**, 575–580 (1998).
8. Israel, M., Meunier, F. M., Morel, N. & Lesbats, B. *J. Neurochem.* **49**, 975–982 (1987).
9. Almers, W. & Tse, F. W. *Neuron* **4**, 813–818 (1990).
10. Choi, S., Klingauf, J. & Tsien, R. W. *Nature Neurosci.* **3**, 330–336 (2000).
11. Wilkens, S., Vasilyeva, E. & Forgacs, M. *J. Biol. Chem.* **274**, 31804–31810 (1999).

Fluid dynamics

That sinking feeling

Michael P. Brenner and Peter J. Mucha

The physics of slowly falling particles in a fluid remains surprisingly enigmatic. Luckily, laws that work for dilute suspensions also appear to apply to higher — and more useful — particle concentrations.

Particles moving through a viscous fluid interact with each other, because each individual particle drags fluid along with it, which then drags on other particles. These hydrodynamic interactions are important in many processes, from the deposition of silt in a river to the motion of dust in the air and the centrifugation of proteins. But despite a century of research, the character of these motions is still hotly debated, even for dilute suspensions in which the particles are well separated. On page 594 of this issue, Segrè *et al.*¹ provide experimental evidence that the laws governing particles settling under gravity at high concentrations are quantitatively similar to those at very low concentrations. This result is unexpected, because it has long been assumed

that particle motion in a dense suspension is qualitatively different from the dilute case.

Smoluchowski carried out an early study of particle interactions in viscous fluids, described in a 1912 paper², “On the practical applicability of Stokes’ law ...”. Smoluchowski wanted to calculate the average sedimentation velocity of identical solid particles, each of which is heavier than the underlying fluid, as they slowly sink through the fluid. The calculation of the average velocity has been revised every 30 years since — by Burgers³ in 1942 and Batchelor⁴ in 1972. Batchelor’s calculation works at low particle concentrations, and assumes that the particle positions are sufficiently random. Experiments have generally shown that Batchelor’s prediction

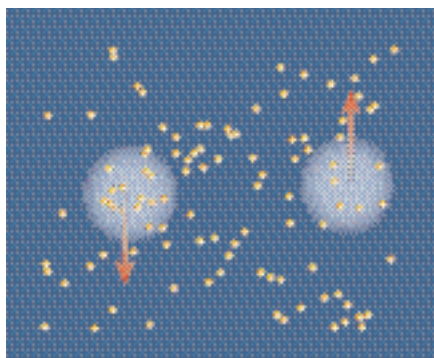


Figure 1 The physical mechanism underlying velocity fluctuations of particles in a sediment. A random distribution of particles has particle-rich regions, which fall faster than the average, and particle-depleted regions, which fall more slowly and so rise relative to the average. A fundamental question in sedimentation research is what sets the size of the particle-rich and particle-depleted regions. (Adapted from an argument by E. J. Hinch.)

is consistent with the data at low particle concentrations (for examples, see ref. 5). But at higher concentrations, strong interactions between groups of particles produce significant complications, which can be understood using sophisticated computer simulations⁶.

However, the average velocity does not completely characterize the sedimentation process — individual particle motions fluctuate about the average. Rather curiously, little effort appears to have been directed at understanding these fluctuations until 1985, when Caflisch and Luke⁷ pointed out that Batchelor's assumptions implied that the variance of particle velocities about the mean scaled linearly with the size of the box containing the sediment — the bigger the box, the bigger the velocity fluctuations. This result is counterintuitive, flying in the face of common experience with normal materials — it implies that the diffusivity of the sedimenting particles depends on the size of the container. In contrast, we know that the diffusion of dye molecules moving in a cup of liquid is controlled by the temperature, the viscosity of the liquid, and the size of the dye molecules, all of which are local quantities and independent of the size of the cup.

Although numerical simulations⁸ support the Caflisch–Luke idea, experiments^{9,10} have indicated that the velocity fluctuations are independent of the box size. In particular, experiments with dilute solutions¹¹ appear to indicate that ‘universal’ concentration-dependent laws govern the velocity fluctuations, independent of the container shape. In these experiments the volume fraction of solid particles varies between 0.01 and 5%. The interpretation of these experiments, and the mechanisms determining the size of the velocity fluctuations, has generated a lot of controversy.

Segrè *et al.*¹ now take these experiments a

step further, by demonstrating that the sedimentation laws described in ref. 11 also apply to much higher volume fractions (between 5 and 50%), as long as three well-studied physical effects are taken into account. These are the dependence of the average sedimentation velocity on concentration; the dependence of the effective viscosity on concentration; and the structure factor, which describes the limitations imposed on otherwise random particle positions as the volume fraction increases and the particles become both more closely packed and more ordered. Segrè *et al.* show that, by correcting the dilute sedimentation laws¹¹ for these effects, they can predict velocity fluctuations across the full range of volume fractions.

Their result contradicts the conventional wisdom that dense suspensions involve very different physics from their dilute counterparts. At high volume fractions the particles are closer together, so particle jamming and other subtle hydrodynamic effects should be more important. In dilute suspensions, well-separated particles interact most strongly through long-range flows generated by each moving particle; but in dense suspensions, particle motion is inhibited by particles that are almost touching. Because most practical applications occur at high concentrations⁵, research into dilute sedimentation has sometimes been dismissed as the study of ‘contaminated water’ — of intellectual interest, but irrelevant for practical problems. The work of Segrè *et al.* shows that the physics of the two regimes are fundamentally connected.

Segrè and colleagues have raised the stakes for figuring out what controls the size of the velocity fluctuations in sedimentation, as we now know that the same theory will apply to all particle concentrations. The physical mechanism that sets the size of these velocity fluctuations is still very much an open question (Fig. 1). In their paper, Segrè *et al.* correct the dilute scaling laws of ref. 11 to obtain predictions of how velocity fluctuations vary with volume fraction — but a similar correction to the Caflisch–Luke argument could also be fit to their data. This brings us back to the fundamental diffusion confusion: is the diffusivity of a sediment a local quantity, as in the case of a dye molecule moving in water, or does it depend on the size of the container?

Here, Segrè *et al.* identify a natural ‘effective temperature’ for sedimentation, which depends on the change in gravitational energy caused by the movement of particle-rich and particle-depleted regions. This is not a real temperature, but a concept borrowed from statistical mechanics, in which a ‘characteristic temperature’ describes the random behaviour in the system. The physical mechanisms setting the size of the fluctuations in particle number — and therefore the effective temperature, velocity fluctuations, and diffusion constant — are due for another revision. Perhaps 30 years after



100 YEARS AGO

The current number of the *Proceedings of the Royal Society* contains a paper of much interest to all who are devoted to the canine race. It describes Dr. Copeman's successful endeavours to isolate the micro-organism responsible for distemper in dogs... Dr. Copeman has now isolated a small *coccobacillus*, growing readily on most of the ordinary culture media at the body temperature, from the exudations from the lungs, the tracheal mucus, and from the nasal secretion of dogs suffering from distemper. A cubic centimetre of a broth-culture of this microbe, injected beneath the skin of the abdomen in a dog weighing 7 kilograms, is sufficient to induce an attack of distemper terminating fatally in about a week from the date of inoculation. A vaccine has also been prepared which Dr. Copeman states can protect dogs against attacks of distemper... An injection of 2 cubic centimetres of such vaccine was apparently sufficient to protect fox-terrier pups weighing about 1½ kilograms when exposed to distemper infection. From *Nature* 31 January 1901.

50 YEARS AGO

Attempts have been made to determine whether pure tones of equal loudness but different pitch are also different in quality, and whether it is possible to find frequency differential thresholds for the quality differences. Following the usual practice, the terms ‘loudness’, ‘pitch’ and ‘quality’ are used to refer specifically to auditory sensations, while ‘intensity’ and ‘frequency’ refer to characteristics of the sound wave stimulus. In this preliminary work, the experimenter was the subject for the comparisons, which were made monaurally with an earphone. It was found possible not only to observe quality differences, but also to describe them. The descriptions were always metaphorical and never altogether satisfactory but appeared to be fairly stable, that is, they seemed equally apt even after long periods of practice in listening and long rest periods over the course of several months. The following are typical of the listener's remarks: 500 c.p.s. sounds ‘purer, clearer, more bell-like’ than 125 c.p.s.; 250 c.p.s. sounds ‘somewhat narrower, thinner and not so fuzzy’ as 125 c.p.s.; 2,000 c.p.s. sounds ‘sharp and hard’ when compared with the ‘soft, blunt, more comfortable’ 125 c.p.s.; 1,500 c.p.s. has a ‘reedy’ quality not observed in 1,000 c.p.s. From *Nature* 3 February 1951.

Batchelor's calculations, we can hope for a new theory in 2002.

Michael P. Brenner and Peter J. Mucha are in the Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307, USA.
e-mails: brenner@math.mit.edu
much@math.mit.edu

1. Segrè, P. N., Liu, E., Umbanhowar, P. & Weitz, D. A. *Nature* **409**, 594–597 (2001).

2. Smoluchowski, M. *Proc. 5th Int. Congr. Math.* **2**, 192–201 (1912).
3. Burgers, J. M. *Proc. Ned. Akad. Wetensch.* **45**, 126–128 (1942).
4. Batchelor, G. I. *J. Fluid Mech.* **52**, 245–268 (1972).
5. Davis, R. H. & Acrivos, A. *Annu. Rev. Fluid Mech.* **17**, 91–118 (1985).
6. Brady, J. F. & Bossis, G. *Annu. Rev. Fluid Mech.* **20**, 111–157 (1988).
7. Caflisch, R. E. & Luke, J. H. C. *Phys. Fluids* **28**, 759–760 (1985).
8. Ladd, A. J. C. *Phys. Fluids* **9**, 491–499 (1997).
9. Ham, J. M. & Homsy, G. M. *Int. J. Multiphase Flow* **14**, 533–546 (1988).
10. Nicolai, H. & Guazzelli, E. *Phys. Fluids* **7**, 3–5 (1995).
11. Segrè, P. N., Herzog, E. & Chaikin, P. N. *Phys. Rev. Lett.* **79**, 2574–2577 (1997).

Functional genomics

Silent genes given voice

Athel Cornish-Bowden and María Luz Cárdenas

Many genes have little apparent influence on growth rates or metabolic fluxes in an organism. But their roles can be revealed by comparing the effects of mutations on two or more metabolite concentrations.

Most mutations have no noticeable impact on an organism. This implies that changing the activity of some enzyme or other by a substantial factor has little effect; even complete deletion of a gene may not be easily detectable if there are appropriate fail-safe features in the design of the organism. Many genes, up to 85% of those in yeast, do not appear to be required for survival, and a high proportion of these seem to have no detectable effects on metabolic fluxes — the chemical processes that result in energy production or growth. This presents a major barrier to functional studies of a genome. How can we hope to deduce the function of a gene that has no apparent effect?

Writing in *Nature Biotechnology*, Léonie Raamsdonk and colleagues argue that examining metabolite concentrations, which in total are known as the 'metabolome', rather than fluxes, is much more likely to reveal such 'silent' genes (*Nature Biotechnol.* **19**, 45–50; 2001). The authors call their method FANCY, which comes from 'functional analysis by co-responses in yeast' (in this case brewer's yeast, *Saccharomyces cerevisiae*). It uses the fact, long known but often ignored, that typical effects of changes in enzyme activity on metabolite concentrations are much larger than their effects on metabolic fluxes. The authors looked at two mutations affecting 6-phosphofructo-2-kinase, an enzyme whose

product fructose 2,6-bisphosphate acts as a signal to regulate energy production; two mutations in the cytochrome oxidase complex, which catalyses a different energy-related process; and a mutation not related to energy metabolism that was used as a control.

Why, though, should changes in enzyme activity have a larger effect on metabolite concentrations than on fluxes? The reason can be seen by looking at what happens when a rock falls into a river. Any transient interruption of the flow of water is rapidly nullified by the increasing level of water just above the rock and the decreasing level just below: as soon as the required pressure is reached, the flow returns to just what it was before. So the rock has no steady-state effect on the flow, although it does have a long-term effect on the water levels, which remain different from their original values as long as the obstacle remains in place. An observer with access only to the steady-state value of the flow can learn nothing about the existence of an obstacle, let alone its location. But an observer with access to the water levels in different places can both detect that an obstacle is present and find out where it is by comparing its effects on the levels above and below it.

As with rivers, so with the genome of an organism such as yeast. Gross properties such as growth rate that depend on fluxes may suggest that most genes are silent. For instance, chemostat culture allows microorganisms to be maintained indefinitely in the phase of exponential growth in a medium of constant composition, the slower-growing strains gradually being eliminated by dilution. Although this approach can in principle reveal very slight differences in growth rates, Raamsdonk *et al.* found that it failed to reveal the deletion of one or other of the two yeast genes that code for 6-phosphofructo-2-kinase. The mutants achieve growth rates equal to those of the wild type by increasing the concentrations of metabolites upstream from the impediment and decreasing those of downstream metabolites (just as for the rock in the river). The effects on a metabolite such as fructose 6-phosphate are not only easily measurable but also detectably different for the two mutants.

Measurements of relatively few metabolite concentrations can thus give voice to apparently silent genes. But individual concentrations are often less informative than one might wish, because quite different mutations may affect the same concentration to a similar extent; in any case, with a completely unknown gene there is no prior knowledge of which concentrations to examine. What is needed is a comprehensive way of studying many metabolites together. The FANCY method provides this, and can reveal quite subtle effects of changes in genotype. The C in FANCY stands for 'co-response analysis' (Hofmeyr, J.-H. S. & Cornish-Bowden, A. *J. Theor. Biol.* **182**, 371–

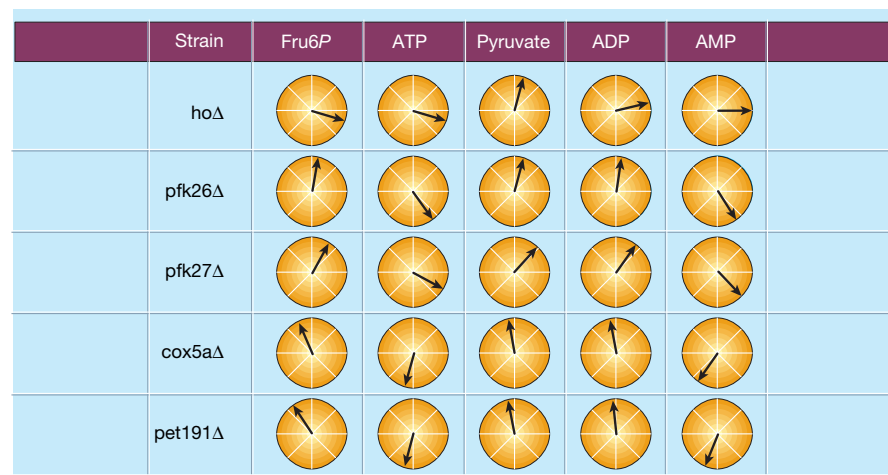


Figure 1 Study of silent genes using co-response angles, from the experiments described by Raamsdonk *et al.* Five yeast strains are shown (Δ indicates a mutant in which a gene has been deleted). The *pfk26Δ* and *pfk27Δ* mutants are defective in 6-phosphofructo-2-kinase activity; *cox5aΔ* and *pet191Δ* are mutants of the cytochrome oxidase complex; *hoΔ*, included as a control, is defective in a part of metabolism not concerned with energy production. Changes in all of the metabolite concentrations shown are relative to changes in the glucose 6-phosphate concentration, which corresponds to the horizontal axis in all cases. The other metabolites are fructose 6-phosphate (Fru6P); pyruvate; and the tri-, di- and monophosphate forms of adenosine phosphate. ATP is the primary energy carrier in the cell, and all the metabolites shown are involved in energy production. Note that the different classes of mutants are characterized by different angles. For simplicity, only the angles are illustrated here, but a full analysis would take account of the magnitude as well as the direction of each vector.

380; 1996). Raamsdonk and colleagues' paper shows how powerful the approach can be.

When two metabolic variables (not necessarily concentrations, but these are often the most revealing) both respond to a change in conditions, their combined response has a direction that can be expressed as an angle. This is better than the corresponding ratio, which suffers from severe problems of statistical instability. The angle is 0° if one does not change at all, 90° if the other does not change at all, 45° if they change by the same amounts in the same direction, and so on. Raamsdonk *et al.* use angles from -90° to $+90^\circ$, but the signs of the individual concentration changes allow use of the full range from -180° to $+180^\circ$, with greater sensitivity.

In Fig. 1 the two 6-phosphofructo-2-kinase mutants (pfk26 Δ and pfk27 Δ) are characterized by similar angles in five different comparisons, easily distinguishable from the angles shown by the other mutants. The two mutants impaired in a different part of energy metabolism (cox5a Δ and pet191 Δ) show angles that are similar to one another but different to those of the other mutants. So we can conclude that co-response analysis does provide a sensitive way of grouping mutations into functional classes.

Co-response analysis can classify even

completely unknown genes into related groups. Combined with multivariate statistical techniques, such as principal-components and discriminant-function analysis, it allows a large amount of metabolite concentration information, from the nuclear magnetic resonance spectra from different mutants for example, to be put into a map where mutants of related function cluster together. Now that this approach has been shown to perform well on questions where the correct answers are already known, the next step is to begin applying it to the thousands of genes in yeast with no known functions — and that is what Raamsdonk *et al.* are now doing. The problem of how to analyse mountains of data arises in all branches of functional genomics. Two other levels of analysis — the proteome (an organism's protein complement) and transcriptome (its messenger RNAs) — may benefit from similar approaches to those being developed for the metabolome.

Athel Cornish-Bowden and María Luz Cárdenas are in the Laboratoire Bioénergétique et Ingénierie des Protéines, Centre National de la Recherche Scientifique, 31 chemin Joseph-Aiguier, BP 71, 13402 Marseille Cedex 20, France.
e-mails: athel@ibsm.cnrs-mrs.fr
cardenas@ibsm.cnrs-mrs.fr

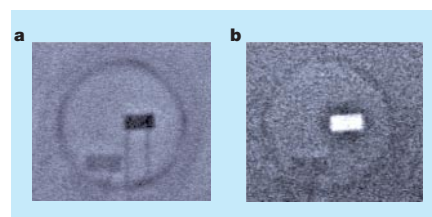


Figure 1 Cooling a solid with light. Thermal images of thulium-doped ZBLANP glass recorded by Hoyt *et al.*². Bright areas indicate cooling, dark areas indicate heating. a, The laser heats the sample as normal. b, The sample is cooled by a laser beam at the right wavelength. (There is some residual heating of the round sample chamber and a reference sample in the bottom left corner.)

and that emitted is caused by a thermal relaxation process that generates a small amount of heat. So even a material that luminesces 99% of the time will still heat up when illuminated.

The key to creating a cooling effect is a process known as 'anti-Stokes luminescence'. This occurs when a material emits light at a higher energy than that used to excite it, because it has excited atoms or ions that have a slight excess of thermal energy in the ground state. These are often referred to as hot or active species. In their experiment, Hoyt *et al.*² use a heavy-metal fluoride glass called ZBLANP (ZrF₄-BaF₂-LaF₃-AlF₃-NaF-PbF₂), doped with thulium ions as the active species. When the hot ions absorb light they retain no memory of their original excited state and so can luminesce to an even higher energy, leaving the ions with less thermal energy than when they started. Repeating this process removes thermal energy from the system, lowering the temperature (Fig. 1). A laser is needed to provide intense light at the correct wavelength for cooling, and the right material is needed to achieve anti-Stokes photoluminescence with an efficiency close to 100%. For their specific system, Hoyt *et al.* report a photoluminescence quantum yield of 99%.

Laser-induced cooling was first observed in a sample of the same glass, ZBLANP, but doped with a different lanthanide ion, ytterbium³. In later experiments⁴, ytterbium-doped glass was laser-cooled to a record 65 °C below room temperature. But by using thulium as the active species, Hoyt *et al.*² shift the wavelength of the laser light from the near-infrared region of the electromagnetic spectrum into the mid-infrared, thereby doubling the intrinsic efficiency of the cooling process. At this wavelength, they cooled a sample of thulium-doped glass to 1.2 °C below room temperature with 40 milliwatts of laser light, corresponding to a cooling efficiency of -30°C per watt of absorbed laser power. Laser cooling has also previously been observed for fluorescent, molecular dyes dissolved in fluid solution^{5,6}, despite the laser wavelength being shifted in the opposite direction, from the infrared into the visible. So laser cooling has now

Solid-state optics

A laser that turns down the heat

Garry Rumbles

An intense laser beam might be expected to cut, burn or blast anything in its path. But at the right wavelength and with a suitable target material, laser light can also chill.

It seems counterintuitive to use laser light to cool a solid, because lasers are commonly perceived to burn. The phaser weapons from science fiction and the lasers used to cut metals both heat their targets to a critical, and sometimes fatal, temperature. But laser light is now routinely used in the laboratory to cool dilute gases or vapours. In some cases the atoms or ions can be cooled to temperatures close to absolute zero, forming a state of matter known as a Bose-Einstein condensate¹. In contrast, techniques to cool solids with light have developed much more slowly. Writing in *Physical Review Letters*, Hoyt and colleagues² claim to have cooled a sample of solid glass with surprisingly high efficiency, by using infrared laser light. This is only the second type of solid to be cooled by lasers.

When an atom absorbs the energy in a light beam, an excited state is created as an electron is promoted from a low-energy to a high-energy state. In photosynthesis, for example, this extra energy is used to convert carbon dioxide and water to carbohydrates and oxygen. Some chemical reactions can

also be used to harness this energy (photochemistry), but in the main it is released either as light (luminescence) or after being converted to thermal energy (heat). When the light source is an intense laser, it is this conversion to heat that leads to rapid temperature rises in the sample. If the target is a flammable solid, it burns; if it is a liquid, it boils.

Alternatively, luminescence can bypass the generation of excess thermal energy and so reduce sample heating. To minimize heating, the ratio of luminescence to heat generation, known as the photoluminescence quantum yield, must be as close to 100% as possible. This is not quite the end of the story, however. Almost all materials that luminesce do so at an energy that is lower than the energy they absorbed — the light emission is said to be 'Stokes shifted' to a lower energy, and hence a longer wavelength. This effect can be observed by using ultraviolet light to irradiate a solution of tonic water containing the molecule quinine. Under these conditions, the solution emits a beautiful turquoise light. The small difference in energy between the light absorbed

been observed in the gas, liquid and solid phases.

Cooling by anti-Stokes luminescence is a thermodynamically spontaneous process. As the temperature of the system decreases, the second law of thermodynamics tells us that the entropy (disorder) of the surroundings must increase by a greater amount. This excess entropy can be found in the emitted luminescence, which occurs in all spatial directions and over a broad spectrum of wavelengths, unlike the laser beam. Landau resolved this issue as long ago as 1946, following the initial proposal of optical cooling by Pringsheim in 1929. Although the experimental verification of these ideas has taken some time, it is unlikely that the commercialization of the idea will take as long.

Indeed, the possibility of manufacturing an all-optical, solid-state cooling device has already been recognized by Ball Aerospace and Technologies Corp. (Colorado, USA) and Metrolaser Inc. (California, USA). Both companies are developing optical refrigerators based on anti-Stokes photoluminescence. Such devices have no moving parts

and are noiseless, which explains their attraction for space and other optoelectronic applications. Metrolaser is developing fluorescent dyes as the active cooling medium, whereas Ball Aerospace is using the original ytterbium-doped ZBLANP system. Lanthanide-doped glasses are already available as optical fibres, so they provide an excellent source of high-quality materials for developing optical refrigerators. It won't be long before thulium-doped glass attracts similar commercial interest.

Garry Rumbles is at the Centre for Electronic Materials and Devices, Department of Chemistry, Imperial College, London SW7 2AY, UK, and the Center for Basic Science, National Renewable Energy Laboratory, Golden, Colorado 80401-3393, USA. e-mail: garry_rumbles@nrel.gov

1. Anderson, M. H. *et al. Science* **269**, 198–201 (1995).
2. Hoyt, C. W., Sheik-Bahae, M., Epstein, R. I., Edwards, B. C. & Anderson, J. E. *Phys. Rev. Lett.* **85**, 3600–3603 (2000).
3. Epstein, R. I., Buchwald, M. I., Edwards, B. C., Gosnell, T. R. & Mungan, C. E. *Nature* **377**, 500–503 (1995).
4. Gosnell, T. R. *Opt. Lett.* **24**, 1041–1043 (1999).
5. Zander, C. & Drexhage, K. H. in *Advances in Photochemistry* Vol. 20 (eds Neckers, D. C., Volman, D. H. & Von Büna, G.) 59–78 (Wiley, New York, 1995).
6. Clark, J. L. & Rumbles, G. *Phys. Rev. Lett.* **76**, 2037–2040 (1996).

Structural biology

Pumping DNA

Edward H. Egelman

Crystal structures of proteins not only shed light on how those proteins work. By revealing previously hidden similarities, they can also force a re-evaluation of what other proteins are predicted to do.

Atomic structures of proteins often reveal evolutionary relationships between them that cannot be seen when looking at sequences of amino acids. A particular fold or topology might be preserved in two proteins from different species, or in different proteins from the same species, even though their amino-acid sequences have diverged so far during evolution that they have no recognizable similarities. This conservation of structure reinforces the idea that there is

only a limited number of stable shapes that proteins can have. It also shows how studies of bacteria or viruses can provide an insight into some processes in higher organisms.

On page 637 of this issue¹, Gomis-Rüth and colleagues show that a protein responsible for transferring single-stranded DNA between bacteria is part of a huge family of proteins involved in the replication, repair, recombination and expression of DNA, as well as in transforming energy. The discovery

provides insight into how this protein might function as a DNA pump. It also tells us more about this ubiquitous protein superfamily.

The first member of the superfamily to have its structure solved was the protein RecA from the bacterium *Escherichia coli*². This intensively studied enzyme is involved in homologous genetic recombination — the process by which two DNA molecules with similar sequences swap strands of DNA. Recombination is a way of repairing DNA, maintaining the stability of the genome, and generating genetic diversity. (In higher organisms, recombination occurs during the production of reproductive cells.) The main active form of RecA appears to be a helical filament formed on DNA, but the protein also forms rings containing six RecA proteins³.

The next protein to be included in this superfamily was a helicase: despite having very different amino-acid sequences, RecA and this helicase⁴ have the same core. Helicases use energy derived from the hydrolysis of ATP to separate double-stranded DNA or RNA into two single strands. These proteins were originally studied in the context of DNA replication — double-stranded DNA must be separated into two strands to allow access by the replicating machinery. But it is now known that helicases participate in all aspects of DNA metabolism, including recombination, replication, repair and transcription. An analysis of the genome of the yeast *Saccharomyces cerevisiae* showed that about 2% of its genes (at least 134 genes) encode helicases, as judged by the presence of certain sequence motifs⁵.

Subsequent structural studies have provided support for the prediction⁴ that all helicases contain the same core as RecA. And just as RecA can exist in a ring form, so too can some of the most well characterized helicases⁶. There is a growing consensus that many of these function by allowing one strand of DNA to pass through the centre of the ring while the complementary strand is displaced outside the ring⁷. However, a bacterial helicase called RuvB appears to function in recombination by pumping

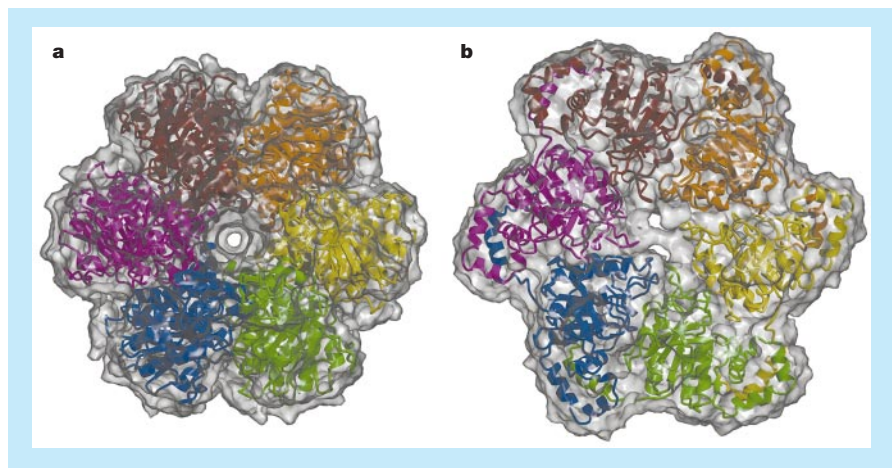


Figure 1 Conservation of structural organization in proteins with very different amino-acid sequences. a, The hexamer formed by the bacterial conjugation protein TrwB, the structure of which has just been determined by Gomis-Rüth and colleagues¹. b, The hexamer formed by the replication helicase gp4 (ref. 6) from a bacterial virus, T7. The structures shown are fragments of the full-length proteins. The part of the TrwB protein shown appears to pump single-stranded DNA, which passes through the central channel. The replication helicase likewise pumps single-stranded DNA through the central channel and so separates the two strands of double-stranded DNA. Figure courtesy of M. Coll.

double-stranded DNA through the central channel of its hexameric ring⁸.

Another addition to this protein superfamily came as a greater surprise. The core of RecA and many helicases is also seen in the F_1 -ATPase⁹. This protein, found only in higher organisms, is part of the F_1F_0 -ATP synthase, which makes ATP — the main source of energy for most cellular processes. So it seems that although the functions (and sequences) of proteins such as helicases and the F_1 -ATPase have diverged greatly, aspects of their organization into hexameric rings have been preserved.

Gomis-Rüth *et al.*¹ now add the *E. coli* protein TrwB to the family. They have determined the structure of TrwB, and find that it exists as a hexameric ring that is remarkably similar to those formed by the F_1 -ATPase, RecA and certain helicases (Fig. 1). The structure and topology of the core regions are also very similar.

The TrwB protein is embedded within the bacterium's outer membrane and is encoded in a small, circular DNA molecule that is not part of the main chromosome. TrwB is needed for conjugation — the transfer of single-stranded DNA between bacterial cells. This process is crucial for the spread of antibiotic resistance within a population of bacteria. Gomis-Rüth *et al.*'s crystal structure¹ did not contain DNA, but a mechanism can still be put forward for how the TrwB ring transfers it between bacteria. TrwB requires energy from ATP hydrolysis to allow conjugation to proceed, so it seemed reasonable that TrwB might be acting as some form of energy-driven motor to move DNA between cells. It now seems likely that the protein not only forms the transmembrane pore that is needed for such transfer, but also acts as the ATP-driven pump that transfers the DNA, working in a similar way to RuvB, mentioned above⁸.

Gomis-Rüth *et al.*'s results¹ also have more general implications. Many proteins have been predicted to be helicases because they contain sequence motifs that are now understood to be part of the conserved core. But neither TrwB nor RuvB works by separating strands of DNA, so these predictions may be incorrect. The putative helicases might be DNA motors involved in pumping or moving along DNA.

Another implication is that proteins that are not classified as helicases, such as the protein SpoIIIE from the bacterium *Bacillus subtilis*, may also be part of this structurally similar superfamily. SpoIIIE is involved in transferring DNA between a *B. subtilis* mother cell and a prespore. This protein actually has a similar amino-acid sequence to that of TrwB, and is capable *in vitro* of moving along DNA in an ATP-dependent manner¹⁰. Perhaps SpoIIIE also forms a hexameric ring around DNA, functioning *in vivo* as yet another DNA pump.

Further structural studies of such pro-

teins will be essential in working out how cells move DNA around. They may also tell us about the evolutionary history of some of the basic machinery for transforming energy and manipulating DNA.

Edward H. Egelman is in the Department of Biochemistry and Molecular Genetics, University of Virginia Health Sciences Center, Charlottesville, Virginia 22908, USA.

Materials science

Ultrafast colour displays

Michael Grätzel

Materials that change their colour as a result of a simple electric potential could be key to a new generation of flat-screen displays. But the speed at which they undergo this change of hue has held them back, until now.

Electrochromic devices are transparent materials that become coloured when an electric potential difference is applied to them. This useful property has already been exploited commercially — for example, in car rear-view mirrors that can be darkened to stop lights from other vehicles from blinding the driver. But the full potential of electrochromic devices has yet to be realized. They could appear as electrochromic windows — panes of glass that can be darkened at the flick of a switch to provide privacy. Alternatively, they could be used in high-contrast dynamic displays such as flat-panel television and computer screens. The main reason these applications remain on the drawing board is the long time it takes present electrochromic devices to switch between transparent and coloured states.

Writing in the *Journal of Physical Chemistry B*, Cummins *et al.*¹ report a new type of

e-mail: egelman@virginia.edu

1. Gomis-Rüth, F. X. *et al.* *Nature* **409**, 637–641 (2001).
2. Story, R. M., Weber, I. T. & Steitz, T. A. *Nature* **355**, 318–325 (1992).
3. Yu, X. & Egelman, E. H. *Nature Struct. Biol.* **4**, 101–104 (1997).
4. Subramanya, H. S. *et al.* *Nature* **384**, 379–383 (1996).
5. Shiratori, A. *et al.* *Yeast* **15**, 219–253 (1999).
6. Singleton, M. R. *et al.* *Cell* **101**, 589–600 (2000).
7. Yu, X. *et al.* *Nature Struct. Biol.* **3**, 740–743 (1996).
8. Parsons, C. A. *et al.* *Nature* **374**, 375–378 (1995).
9. Abrahams, J. P. *et al.* *Nature* **370**, 621–628 (1994).
10. Bath, J. *et al.* *Science* **290**, 995–997 (2000).

electrochromic system that promises to overcome this problem. Their device yields high-contrast images that can be rapidly created and erased. In addition, the optical appearance of the images does not depend on the angle from which they are viewed, an important attribute for use in display screens.

There are two basic types of electrochromic device in use at the moment. One uses a thin, transparent metal-oxide film, such as tungsten trioxide (WO_3), which is deposited onto a conducting glass substrate. When a negative potential is applied to the glass, there is an influx of conducting electrons into the metal-oxide film, which turns the WO_3 blue. Because the system wants to be electrically neutral, this causes positive ions, usually lithium (Li^+), to move into the film from the surrounding electrolyte. Without this charge compensation, the electrons could not be injected into the film. But the

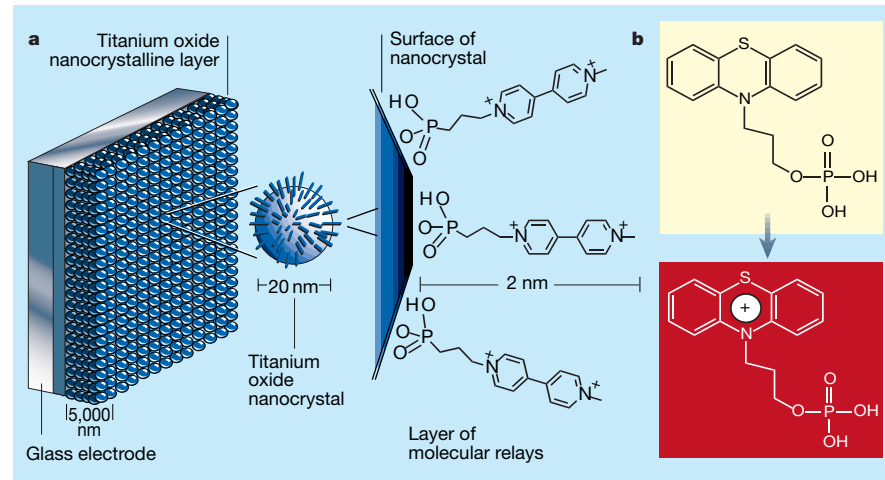


Figure 1 Ingredients for making a dynamic colour display. Cummins *et al.*¹ have built an electrochromic cell from two porous metal-oxide films, sandwiched between glass electrodes. a, The negative electrode is coated with a layer of TiO_2 nanocrystals, with viologen molecules anchored to the surface of the nanocrystals. Viologen molecules turn blue when injected with charge. b, The positive electrode consists of a nanocrystalline $SnO_2:Sb$ film, which is linked to phenothiazine molecules, which turn red when oxidized. When the electrochromic cell is filled with an electrolyte and a voltage is applied, the colour can be switched on and off in under 250 milliseconds.

diffusion of the positive ions into the oxide layer is slow, taking minutes to complete.

Slow charge diffusion also limits the switching speed of the second type of electrochromic device. Here the colour change arises from the oxidation (removal of electrons) or reduction (addition of electrons) of molecules known as 'molecular relays'. A typical example is methyl viologen, which is not coloured in its normal oxidation state, but turns deep blue when reduced. Such relays have so far been used either in solution or incorporated into a polymeric film.

Attempts have been made to overcome the charge diffusion problem by attaching the relays to the rough surface of a porous oxide film²⁻⁴. Typically, the film consists of a network of interconnected semiconducting or conducting metal-oxide nanocrystals, deposited onto a conducting glass electrode. The particles are fused together by heating to form a network that allows electric charges to move through it easily. Because the oxide crystals are so small, the film has an extraordinarily large internal surface area. The ratio between the internal surface area and the electrode's geometrical area (its 'roughness factor') approaches 1,000 for a film that is only a few micrometres thick. This means that a high number of molecular relays can be contained in a relatively small area.

Cummins *et al.*¹ have developed this concept and assembled an electrochromic cell that uses two metal-oxide films — one at the negative electrode and, unusually, one at the positive electrode. The negative glass electrode is coated with a porous titanium dioxide (TiO₂) film, which is itself coated with a single layer of viologen molecules. These molecular relays are anchored firmly to the surface of the TiO₂ nanocrystals by phosphonate groups (Fig. 1a). The positive glass electrode features another nanocrystalline film, this time made up of antimony-doped tin dioxide (SnO₂:Sb). This film has a layer of phosphonated phenothiazine molecules attached to it. In their normal state, these molecules have a weak yellow colour, but when they are oxidized, they turn red (Fig. 1b). The electrochromic cell is filled with an electrolyte and sealed.

The device can be switched from an uncoloured to a deeply coloured state by applying a potential difference of 1.2 volts to the system. This reduces the viologen, turning it blue, and oxidizes the phenothiazine. The overall result is a blue-red colour (Fig. 2). Similarly, reversing the potential removes the colour from the system. The cycle can be repeated many thousands of times without significant degradation of performance. Moreover, the colour persists for more than 600 seconds after the voltage is switched off.

Several remarkable features make this system very attractive for use in dynamic displays. The huge internal surface area of the two nanocrystalline films makes the electrochromic process very efficient, even though



Figure 2 Prototype of a fast switching, high-contrast electrochromic device viewed at a low angle.

the electron relay forms just a single molecular layer across the surface of the film's porous interior. On a flat surface the optical absorbance (colour change) caused by reduction or oxidation of the relay would be hardly visible. But when deposited on a nanocrystalline film, the coloration of the molecular relays is amplified by more than 1,000 times because of the high number of relay molecules present. The coloration efficiency, expressed as the change in the film's optical absorbance per unit of injected charge, is higher for these systems than for conventional electrochromic layers. This is because an electric charge injected into an organic molecule, such as viologen, which produces intensely coloured radical ions, causes a larger change in optical absorbance than when it is transferred to an oxide such as WO₃.

But the most striking aspect of this work is the fast rate at which the device can be switched. Its speed can be attributed to a combination of rapid charge-motion within the metal oxide, ultrafast electron-transfer from the nanocrystalline metal-oxide to the relay molecule, and efficient movement of the charge-balancing ions in the electrolyte within the pores^{3,4}. The switching times achieved by Cummins *et al.* are in the range of milliseconds, which is about a thousand times shorter than is possible with conventional electrochromic systems. Further

acceleration of the switching cycle into the microsecond range appears possible by judicious choice of the nanocrystalline materials and by reducing the electrical resistance within the electrochromic device.

High-contrast images created in electrochromic devices retain their quality even when viewed at a low angle — in contrast with conventional liquid crystal displays. And with molecular relays there is a wide choice of molecules that can generate a specific colour, unlike devices that use WO₃, which are limited to the colour blue. The colour change depends in an exponential way on the applied potential, allowing the device to be fashioned into a passive matrix display where the surface is divided into pixels. Each pixel can be addressed individually by applying a voltage difference locally at the edge of the panel. The combination of a molecular charge-relay with a porous support opens up exciting new opportunities for low-cost, thin-film molecular electronics. ■

Michael Grätzel is in the Laboratory for Photonics and Interfaces at the Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland.
e-mail: michael.gratzel@epfl.ch

1. Cummins, D. *et al.* *J. Phys. Chem. B* **104**, 11449–11459 (2000).
2. Cinnsealach, R., Boschloo, G., Rao, S. N. & Fitzmaurice, D. *Sol. Energy Mater. Sol. Cells* **57**, 107 (1999).
3. Marguerettaz, X., O'Neill, R. & Fitzmaurice, D. *J. Am. Chem. Soc.* **116**, 2629–2630 (1994).
4. Hagfeldt, A. *et al.* *Proc. SPIE* **2255**, 297–303 (1994).

Oceanography

The Rossby rototiller

David A. Siegel

Planetary waves, also known as Rossby waves, propagate throughout the world's oceans on very large scales. They influence the ocean–climate system and also, it seems, the delivery of nutrients to the ocean surface.

Data from Earth-viewing satellite systems have advanced to the point that marine scientists can carry out global analyses that were inconceivable just a few years ago. An excellent example can be found in the papers by Uz *et al.* on page 597 of this issue¹ and by Cipollini *et al.*² in the 15 January issue of *Geophysical Research Letters*. Both groups find a significant covariation of global chlorophyll concentrations with satellite determinations of sea level. The upshot of their investigations is the recognition that so-called planetary or Rossby waves provide a physical mechanism for bringing nutrients

to the surface layers of the open ocean and feeding phytoplankton productivity.

Like all plants, phytoplankton require both light and nutrients to grow. But only in the upper 50 to 150 m of the water column of the open sea is there enough light for phytoplankton photosynthesis to occur. In this thin, well-lit euphotic zone, available nutrients are rapidly assimilated. Typically, then, it is the supply of new nutrients from deeper water that limits productivity in the open ocean.

In most coastal and high-latitude regions of the sea, vertical mixing and upwelling provide sufficient vertical fluxes of nutrients to

sustain phytoplankton growth. In contrast, the surface layers of much of the subtropical and tropical ocean are oligotrophic (nutrient-poor). In such regions, intense upper-layer stratification through a feature known as the thermocline seals off the euphotic zone from below, inhibiting the upward transport of nutrients. For these regions, diffusion of nutrients by small-scale turbulence is thought to be inadequate to account for observed rates of productivity^{3,4}. So some other physical process must be stirring up the oligotrophic ocean and supplying nutrients to the euphotic zone.

A leading candidate for such a disturbance is 'eddy pumping'^{4,5}. In the most general terms, ocean eddies and meanders on scales of tens to hundreds of kilometres drive the vertical motion of water, and so have important biological and biogeochemical consequences, including lifting subsurface nutrients into the euphotic zone⁴⁻⁸. In the stratified mid-latitude ocean, the formation of cyclonic mesoscale eddies, which are closed vortices roughly 200 km in diameter, are thought to be the dominant physical process driving the eddy-pumping flux^{4,5,9}.

So, is that the end of the eddy pumping story? This is where the work of Uz *et al.*¹ and Cipollini *et al.*² comes in. They find that there is a global statistical coherence between sea level, detected by satellite, and ocean chlorophyll concentration on scales of 300–1,000 km — considerably larger than are typical for mesoscale eddies. Chlorophyll provides a proxy measurement for phytoplankton biomass and hence productivity; sea-level variations primarily reflect changes in the depth of the thermocline. By analysing the spatial scales and propagation velocities of the eddies, the authors interpret these variations as baroclinic planetary waves or Rossby waves. Baroclinic Rossby waves are large-scale (a few hundred to several thousand kilometres across), westward-propagating waves, which create vertical displacements of the thermocline of 10–100 m, and which are driven by the latitudinal change in the Coriolis force¹⁰. They have long been known to be 'messengers' of oceanic climate disturbances. But the ability of satellites to measure sea level to within a few centimetres has added fresh impetus to research into the characteristics and dynamics of these waves^{11,12}.

The demonstration by Uz *et al.* and Cipollini *et al.* of a link between Rossby waves and global chlorophyll distribution has a bearing on the eddy-pumping hypothesis. For example, mesoscale eddies should retain water masses within their core. So nutrients should reach the euphotic zone as the eddy forms, not as it propagates. On the other hand, planetary waves are waves in the classical sense and do not transport water masses with them. So they should be able to lift subsurface nutrients continuously into the euphotic zone as they propagate across

an ocean basin⁹. In a sense, they may act like a rototiller, or rotovator, 'turning over' upper-ocean nutrients in their path. This fundamental difference between waves and eddies will have an important bearing on how physical observations are used to infer eddy-pumping fluxes⁹.

Uz *et al.*¹ find that only 10 to 20% of the variance in chlorophyll (filtered to remove seasonal and other variations) is explained by the sea-level signal. So it seems that the 'Rossby rototiller' may not be the dominant physical process regulating eddy pumping. Further research is also needed to understand how variability in chlorophyll biomass translates to the cycling of carbon and associated materials in the water column. There is a long way to go before a predictive understanding of the interactions of planetary waves and mesoscale eddies on phytoplank-

ton productivity and, more generally, ocean biogeochemical cycles, is established¹³. ■

David A. Siegel is at the Institute for Computational Earth System Science, University of California at Santa Barbara, California 93106-3060, USA.

e-mail: davey@icess.ucsb.edu

1. Uz, B. M., Yoder, J. A. & Osychyn, V. *Nature* **409**, 597–600 (2001).
2. Cipollini, P., Cromwell, D., Challenor, P. G. & Raffaglio, S. *Geophys. Res. Lett.* **28**, 323–326 (2001).
3. Jenkins, W. J. *Nature* **331**, 521–522 (1988).
4. McGillicuddy, D. J. Jr *et al.* *Nature* **394**, 263–266 (1998).
5. McGillicuddy, D. J. Jr & Robinson, A. R. *Deep-Sea Res.* **144**, 1427–1450 (1997).
6. Falkowski, P. G. *et al.* *Nature* **352**, 55–58 (1991).
7. Flierl, G. R. & Davis, C. S. *J. Mar. Res.* **51**, 529–560 (1993).
8. Oschlies, A. & Garçon, V. C. *Nature* **394**, 266–269 (1998).
9. Siegel, D. A., Fields, E. & McGillicuddy, D. J. Jr *J. Geophys. Res.* **104**, 13359–13379 (1999).
10. Gill, A. E. *Atmosphere–Ocean Dynamics* (Academic, New York, 1982).
11. Chelton, D. B. & Schlax, M. G. *Science* **272**, 234–238 (1996).
12. Polito, P. S. & Cornillon, P. J. *Geophys. Res.* **102**, 3215–3235 (1997).
13. Doney, S. C. *Glob. Biogeochem. Cycles* **13**, 705–714 (1999).

Vascular biology

Targeted delivery of nitric oxide

Steven S. Gross

Nitric oxide is a biological signalling gas that has been assumed to reach its protein targets by simple random diffusion. The discovery of molecular mechanisms for precise nitric oxide delivery challenges that assumption.

A major function of nitric oxide (NO) in the body is to regulate blood flow. With every heartbeat, a puff of NO is released from the endothelial cells that line our blood vessels. This NO has been assumed to diffuse passively into the surrounding smooth muscle and to react with proteins that cause blood vessels to relax, thereby increasing blood flow. But, despite universal agreement that the NO comes from endothelial cells, and that smooth muscle is the target, the route and chemical form that NO take in its journey have been controversial. The old assumption that diffusion alone moves NO around has taken a few knocks. A paper on page 622 of this issue¹ deals another blow to the random-diffusion theory. Pawloski and colleagues reveal that the movement of NO through red blood cells occurs by a precisely coordinated series of events. Signalling by NO is not simple, random or mediated exclusively by diffusion.

A big problem with the idea of NO moving by random diffusion is that it reacts rapidly with haemoglobin in red blood cells, suggesting that endothelium-derived NO would be instantly consumed in the blood, leaving none for controlling the blood vessels. This dilemma is not fully resolved by the existence of a red-blood-cell-free zone at the endothelial surface of larger blood vessels² and an unspecified barrier that appears to slow the diffusion of NO through red-blood-cell membranes^{3,4}. One suggested solution is that a physiologically significant fraction of the NO that enters red blood cells — previ-

ously thought to be irreversibly consumed by reactions with haemoglobin — re-emerges and returns to the blood-vessel wall as a biologically active S-nitrosothiol molecule⁵. Here, the S-nitrosothiol acts as a cloaked form of NO activity that is protected from reacting with haemoglobin. Support for this idea was provided by the demonstration that, under certain conditions, blood vessels relax after exposure to red blood cells that have been pretreated with NO (ref. 5).

The discovery that NO's biological activity can be salvaged from red blood cells forced a reconsideration of the reactions between NO and haemoglobin that predominate in physiological settings^{6,7}. Through the wondrous molecular gymnastics of haemoglobin, some of the NO captured by Fe²⁺ in haemoglobin can be shuttled intramolecularly to a conserved thiol group (provided by cysteine residue β 93 in human haemoglobin), producing S-nitroso-haemoglobin. This may be followed by transfer of the NO group to other thiol-containing molecules, which enable NO activity to depart the red blood cell. This would occur preferentially in oxygen-poor tissues, where deoxygenation triggers haemoglobin to change its conformation from the oxygen-bound 'R' structure to the deoxygenated 'T' structure. The beauty of this system is that it provides selective delivery of NO to oxygen-depleted tissues, resulting in blood-vessel relaxation and increased blood flow where it is most needed. But where and how does NO disembark

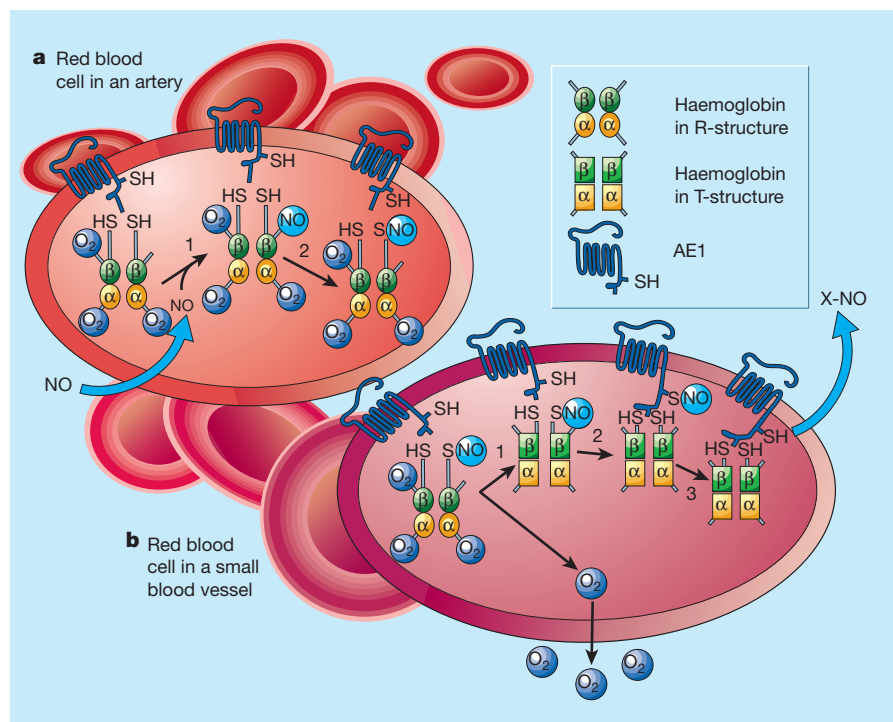


Figure 1 A model for how red blood cells sequentially consume and release nitric oxide (NO) for control of blood flow. **a**, In arteries, the oxygen saturation of blood is high, causing haemoglobin to exist mainly in the R-structure (carrying three to four oxygen molecules). Step 1, NO derived from the blood-vessel lining is efficiently scavenged through reactions with Fe^{2+} in haemoglobin. Step 2, a fraction of scavenged NO can transfer to a thiol (SH) group in haemoglobin, forming SNO-haemoglobin. **b**, In small blood vessels the oxygen saturation of blood is low. Step 1, SNO-haemoglobin releases its oxygen and converts to the T-structure. This promotes the binding of SNO-haemoglobin to the anion-exchange protein AE1 and imposes steric hindrance on bound NO. Step 2, NO is selectively transferred to a thiol in AE1, forming SNO-AE1. Step 3, SNO-AE1 can transfer NO activity out of the red blood cell and into the vessel wall. The mechanism of release and chemical form of 'X-NO' await discovery. These reactions may apply to only 1 in 10,000 haemoglobin molecules that possess NO, but they may be important in regulating blood flow in response to changing oxygen levels.

from S-nitroso-haemoglobin to exit the red blood cell? Pawloski *et al.*¹ present an unexpected first step: targeted delivery of NO to an unlikely new player in NO biology, the anion-exchange protein AE1.

AE1 is the most abundant protein in the plasma membrane of red blood cells — a million copies serve as a scaffold that links a network of cytoskeletal proteins. The recognized biochemical function of AE1 is as an electro-neutral chloride/bicarbonate exchanger that increases the ability of red blood cells to rid themselves of carbon dioxide in the lungs. The protein also anchors a subpopulation of haemoglobin molecules to red-blood-cell membranes, but the functional significance of this has remained obscure. Pawloski *et al.*'s results¹ suggest that this interaction may have evolved specifically to mediate the transfer of NO groups.

The authors loaded red blood cells with physiological amounts of NO *in vitro*, resulting in the incorporation of NO into R-structure haemoglobin. When the cells were then subjected to low-oxygen conditions, haemoglobin converted to its T-structure, and a large fraction of S-nitrosothiol transferred from inside the cell (its cytosol) to its mem-

brane. This redistribution of S-nitrosothiol was prevented by drugs that specifically bind AE1 and inhibit its ability to transport anions. The authors confirmed that S-nitrosothiol groups were in AE1 by precipitating S-nitrosothiol groups using antibodies that recognize AE1. The functional significance of this AE1-mediated transfer of S-nitrosothiol to the membrane is implied by evidence that AE1 inhibitors diminish the vessel-relaxing activity of red blood cells that contain S-nitroso-haemoglobin. So AE1 seems to be the direct and primary recipient of NO groups from S-nitroso-haemoglobin, and this process is essential for the efflux of NO's biological activity from red blood cells (Fig. 1).

There must be exquisite specificity in the transfer of NO from S-nitroso-haemoglobin to AE1. An X-ray crystal structure indicates that the cytosolic part of one end of AE1 docks between the β -subunits of the haemoglobin molecules⁸. Cysteine residue β 93 of haemoglobin is positioned near potential target thiol groups in AE1, as confirmed by Pawloski *et al.*, and two thiols at the cytosolic end of AE1 are associated with amino acids that make up a 'motif' that favours this type of chemical reaction⁹.

Daedalus

David Jones

David Jones, author of the Daedalus column, is indisposed.

Pawloski *et al.* do not identify either the route by which NO activity detaches from S-nitroso-AE1, or the chemical form. An intriguing but unprecedented possibility would be the direct egress of an NO-derived species through the channel of AE1. Other possibilities include the transfer of an NO group (presumably in the form of an S-nitrosothiol) to a molecule that moves to the extracellular face of the plasma membrane and is either released into blood or transfers NO activity to another acceptor. This acceptor could be a component of the endothelial-cell membrane; conceivably, NO activity could be transferred directly as red blood cells bump along the endothelium.

Pawloski *et al.*'s finding that NO groups are transferred directly from protein to protein is unprecedented. This realization emerged from what is arguably the first time a research team has rigorously asked how NO's biological activity leaves a cell. Red blood cells impose extreme hardship for NO efflux, as they contain an enormous concentration of NO-hungry intracellular haem groups. Here, it was essential that molecular strategies evolve for precise NO delivery to avoid relentless capture of NO by haem. Nonetheless, mechanisms for targeted NO delivery have certainly also evolved in more 'typical' NO-producing cell types. For example, NO-producing enzymes are found in some large protein assemblies where NO has been assumed, and recently proven, to be selectively delivered to thiol groups of targets at the same area¹⁰.

New knowledge that NO groups can be transferred with precision to specific protein thiols shows this modification to have broad potential use for moment-to-moment regulation of protein activity. The development of new methods to detect S-nitrosylation that circumvent the problem of the instability of S-nitrosothiol groups will almost certainly affirm this possibility. Simple passive diffusion as the prototypic mode of signalling by NO is a concept that is failing fast.

Steven S. Gross is in the Department of Pharmacology, Cornell University Medical College, 1300 York Avenue, New York, New York 10021, USA. e-mail: ssgross@med.cornell.edu

- Pawloski, J. R., Hess, D. T. & Stamler, J. S. *Nature* **409**, 622–626 (2001).
- Liao, J. C. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 8757–8761 (1999).
- Liu, X. *et al.* *J. Biol. Chem.* **273**, 18709–18713 (1998).
- Vaughn, M. W. *et al.* *J. Biol. Chem.* **275**, 2342–2348 (2000).
- Jia, L. *et al.* *Nature* **380**, 221–226 (1996).
- Gow, A. J. & Stamler, J. S. *Nature* **391**, 169–173 (1998).
- Gross, S. S. & Lane, P. *Proc. Natl Acad. Sci. USA* **96**, 9967–9969 (1999).
- Walder, J. A. *et al.* *J. Biol. Chem.* **259**, 10238–10246 (1984).
- Stamler, J. S. *et al.* *Neuron* **18**, 691–696 (1997).
- Fang, M. *et al.* *Neuron* **28**, 183–193 (2000).

A fern that hyperaccumulates arsenic

A hardy, versatile, fast-growing plant helps to remove arsenic from contaminated soils.

Contamination of soils with arsenic, which is both toxic and carcinogenic, is widespread¹. We have discovered that the fern *Pteris vittata* (brake fern) is extremely efficient in extracting arsenic from soils and translocating it into its above-ground biomass. This plant — which, to our knowledge, is the first known arsenic hyperaccumulator as well as the first fern found to function as a hyperaccumulator² — has many attributes that recommend it for use in the remediation of arsenic-contaminated soils.

We found brake fern growing on a site in Central Florida contaminated with chromated copper arsenate (Fig. 1a). We analysed the fronds of plants growing at the site for total arsenic by graphite furnace atomic absorption spectroscopy. Of 14 plant species studied, only brake fern contained large amounts of arsenic (As; 3,280–4,980 p.p.m.). We collected additional samples of the plant and soil from the contaminated site (18.8–1,603 p.p.m. As) and from an uncontaminated site (0.47–7.56 p.p.m. As). Brake fern extracted arsenic efficiently from these soils into its fronds: plants growing in the contaminated site contained 1,442–7,526 p.p.m. arsenic and those from the uncontaminated site contained 11.8–64.0 p.p.m. These values are much higher than those typical for plants growing in normal soil, which contain less than 3.6 p.p.m. of arsenic³.

As well as being tolerant of soils containing as much as 1,500 p.p.m. arsenic, brake fern can take up large amounts of arsenic into its fronds in a short time (Table 1). Arsenic concentration in fern fronds growing in soil spiked with 1,500 p.p.m. arsenic increased from 29.4 to 15,861 p.p.m. in two weeks. Furthermore, in the same period, ferns growing in soil containing just 6 p.p.m. arsenic accumulated 755 p.p.m. of arsenic in their fronds, a 126-fold enrichment. Arsenic concentrations in brake fern roots were less than 303 p.p.m., whereas those in the fronds reached 7,234 p.p.m.

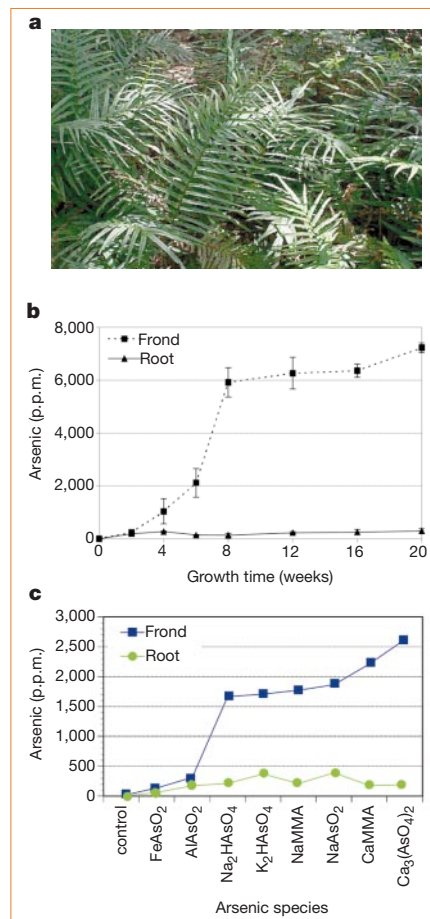


Figure 1 Arsenic concentrations in brake fern growing in arsenic-contaminated soils. **a**, Brake fern growing on an abandoned wood-preservation site contaminated with chromated copper arsenate (CCA); **b**, arsenic concentrations in brake fern after 20 weeks' growth in a CCA soil containing 97 p.p.m. As; and **c**, arsenic concentrations in brake fern after 18 weeks' growth in soil spiked with 50 p.p.m. As of various species. Brake fern plants grown in the laboratory were transferred to 2.5-litre pots (one plant per pot, with four replicates) containing 1.5 kg soil to determine arsenic-uptake changes with time and the arsenic species. NaMMA, monosodium methylarsenate; CaMMA, calcium acid methanearsonate.

(Fig. 1b). Addition of 100 p.p.m. arsenic significantly stimulated fern growth, resulting in a 40% increase in biomass compared

with the control (data not shown).

After 20 weeks of growth, the plant was extracted using a solution of 1:1 methanol:water to speciate arsenic with high-performance liquid chromatography–inductively coupled plasma mass spectrometry. Almost all arsenic was present as relatively toxic inorganic forms, with little detectable organoarsenic species⁴. The concentration of As(III) was greater in the fronds (47–80%) than in the roots (8.3%), indicating that As(V) was converted to As(III) during translocation from roots to fronds.

As well as removing arsenic from soils containing different concentrations of arsenic (Table 1), brake fern also removed arsenic from soils containing different arsenic species (Fig. 1c). Again, up to 93% of the arsenic was concentrated in the fronds. Although both FeAsO₄ and AlAsO₄ are relatively insoluble in soils¹, brake fern hyperaccumulated arsenic derived from these compounds into its fronds (136–315 p.p.m.) at levels 3–6 times greater than soil arsenic.

Brake fern is mesophytic and is widely cultivated and naturalized in many areas with a mild climate. In the United States, it grows in the southeast and in southern California⁵. The fern is versatile and hardy, and prefers sunny (unusual for a fern) and alkaline environments (where arsenic is more available). It has considerable biomass, and is fast growing, easy to propagate, and perennial.

We believe this is the first report of significant arsenic hyperaccumulation by an unmanipulated plant. Brake fern has great potential to remediate arsenic-contaminated soils cheaply and could also aid studies of arsenic uptake, translocation, speciation, distribution and detoxification in plants.

Lena Q. Ma*, Kenneth M. Komar†, Cong Tu*, Weihua Zhang‡, Yong Cai‡, Elizabeth D. Kennelley*

*Soil and Water Science Department, University of Florida, Gainesville, Florida 32611-0290, USA
e-mail: lqma@ufl.edu

†Cooperative Extension Service, University of Georgia, Terrell County, PO Box 271, Dawson, Georgia 31742, USA

‡Department of Chemistry & Southeast Environmental Research Center, Florida International University, Miami, Florida 33199, USA

1. Nriagu, J. O. (ed.) *Arsenic in the Environment Part I: Cycling and Characterization* (Wiley, New York, 1994).
2. Brooks, R. R. (ed.) *Plants that Hyperaccumulate Heavy Metals* (Cambridge Univ. Press, 1998).
3. Kabata-Pendias, A. & Pendias, H. in *Trace Elements in Soils and Plants* 203–209 (CRC, Boca Raton, 1991).
4. Koch, I., Wang, L., Ollson, C. A., Cullen, W. R. & Reimer, K. J. *Envir. Sci. Technol.* **34**, 22–26 (2000).
5. Jones, D. L. *Encyclopaedia of Ferns* (Lothian, Melbourne, 1987).

Table 1 Arsenic concentrations in brake fern

Treatments	Soil arsenic (p.p.m.)	Plant arsenic (p.p.m.)	
		2 weeks	6 weeks
Control	6	755	438
As-contaminated soil*	400	3,525	6,805
Low As†	50	5,131	3,215
Medium As†	500	7,849	21,290
High As†	1,500	15,861	22,630

Brake fern plants, collected from several uncontaminated sites (they are not commercially available), were planted in 2.5-litre pots containing 1.5 kg of soil (one plant per pot with four replicates) and grown for six weeks.

*Arsenic-contaminated soil was collected from the site where brake fern was discovered.

†Artificially contaminated soil was spiked with three levels of water-soluble potassium arsenate.

Neuroperception

Superior auditory spatial tuning in conductors

Conducting a large orchestra is an impressive feat that simultaneously requires the intake of the whole musical *gestalt* and the analytical decomposition of the orchestral sound into its components¹. How, for example, does a conductor identify a specific musician within a multiplayer section? Here we provide evidence from brain-potential recordings that experienced professional conductors develop enhanced auditory localization mechanisms in peripheral space.

Seven classical-music conductors (mean age, 45 yr; mean conducting experience, 19 yr; minimum, 6 yr), seven pianists (mean age, 43 yr; mean professional playing experience, 16 yr; minimum, 7 yr) and seven non-musicians (mean age, 43 yr) were tested in a paradigm used originally to demonstrate superior sound localization in congenitally blind subjects² (Fig. 1a). Specifically, the subjects listened to brief pink-noise bursts delivered by central and peripheral arrays of three loudspeakers each (C1–3 and P1–3 in Fig. 1a); these speakers were arranged along a semicircle extending from the midline to 90 degrees right of centre.

While frequent stimuli (84%; frequency of 500–5,000 Hz, 75 decibels, and 80 ms duration) and infrequent ‘deviant’ stimuli of increased bandwidth (16%; 500–15,000 Hz) were delivered in random order from all speakers (interstimulus interval, 90–270 ms), the subject’s task was selectively to attend — in different runs — to the centre-most (C1) or rightmost speaker (P1) and to press a button to indicate the ‘deviant’ stimuli occurring at the designated location (called targets). We recorded multichannel event-related brain potentials (ERPs) using standard methodology³ and these showed a typically enhanced negativity (termed ‘Nd attention’ effect) for the relevant speakers (Fig. 1a). In the conductors, ERPs invoked in response to stimuli from adjacent locations showed a similar attention effect of smaller amplitude, indicating that auditory spatial attention is distributed in a gradient fashion³ for both central and peripheral auditory space.

Attentional effects are also revealed by computing the difference between ERPs to attended-direction and unattended-direction stimuli (Fig. 1b, c). Although a spatial gradient was evident in all three groups for central auditory space, only the conductors displayed a gradient for the periphery. This improved spatial tuning in conductors also has behavioural consequences, as attested by a significantly reduced false-alarm rate for adjacent locations (P2, P3) in the periphery.

The very similar scalp topography of the

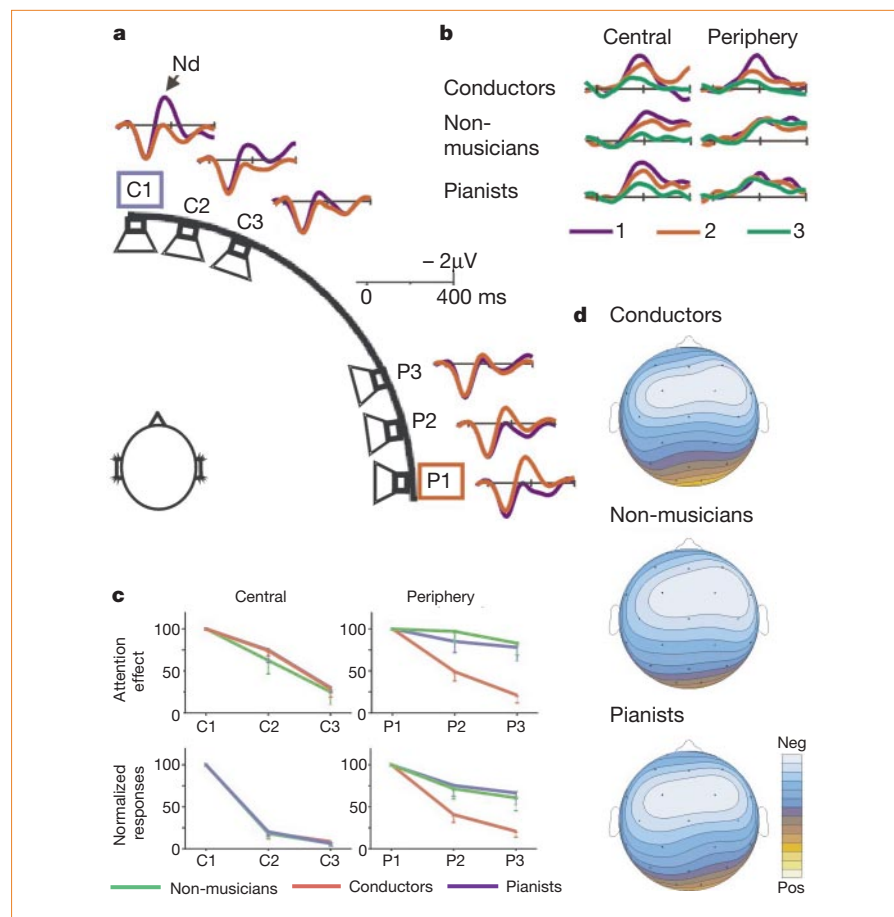


Figure 1 Effects of auditory attention in conductors, pianists and controls. **a**, Experimental set-up; speakers are spaced 6 degrees apart. Group-average event-related potentials (ERPs; frontal midline site) recorded from the conductors and invoked by frequent standard stimuli are represented by blue lines that indicate responses to stimuli from a particular speaker when attending to loudspeaker C1; red lines represent ERPs in response to the same stimuli when attending to speaker P1. Attended stimuli give rise to an enhanced negativity starting at 60 ms (Nd). ERPs associated with adjacent speakers show a similar declining gradient. **b**, Difference waves obtained by subtracting unattended-direction from attended-direction responses. All subject groups show a gradient ERP for central locations; for peripheral sounds, a gradient is evident only for the conductors. **c**, Top row, electrophysiological attention effect (frontal midline electrode; mean amplitude, 180–220 ms; C1/P1 set to 100%). No differences between groups were found for central locations. Conductors showed a steeper gradient in the periphery (groups by location interaction: conductors versus non-musicians, $P=0.015$; conductors versus pianists, $P=0.044$). Bottom row, button presses in response to infrequent stimuli (C1/P1 set to 100%). For peripheral sounds, only conductors show a decreased false-alarm rate for adjacent locations (group by location interaction: conductors versus non-musicians, $P<0.01$; conductors versus pianists, $P<0.01$). **d**, Spline-interpolated scalp maps of the attention effect for the centre-most speaker (time window, 180–220 ms) show a similar topography across groups.

attention effect (Fig. 1d) for the different groups indicates that conductors probably do not engage different neural populations to perform the task. From magnetoencephalographic recordings⁴, the attention effect is known to arise in the secondary auditory cortex, an area also implicated from functional imaging⁵. Improved learning-induced use of spectral cues generated by the head and outer ears, and analysed by the auditory cortex⁶, might underlie the localization advantage experienced by conductors. Although conductors probably employ other mechanisms such as perceptual grouping^{7,8} to identify single musicians, our findings provide another example of how extensive training can shape cognitive processes and their neural underpinnings. **Thomas F. Münte***, **Christine Kohlmetz†‡**, **Wido Nager‡**, **Eckart Altenmüller†**

*Department of Neuropsychology, University of Magdeburg, PO Box 4120, 39016 Magdeburg, Germany
e-mail: thomas.muente@medizin.uni-magdeburg.de
†Institute for Music Physiology and Performing Arts Medicine, Hannover Academy of Music and Theatre, 30175 Hannover, Germany
‡Department of Neurology, Medical School Hannover, 30623 Hannover, Germany

- Schuller, G. *The Complete Conductor* (Oxford Univ. Press, Oxford, 1997).
- Röder, B. et al. *Nature* **400**, 162–166 (1999).
- Mondor, T. A. & Zatorre, R. J. *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 387–409 (1995).
- Woldorff, M. G. et al. *Proc. Natl Acad. Sci. USA* **90**, 8722–8726 (1993).
- Grady, C. L. et al. *NeuroReport* **28**, 2511–2516 (1997).
- Rauschecker, J. P. *Trends Neurosci.* **22**, 74–80 (1999).
- Bregman, A. S. *Auditory Scene Analysis: The Perceptual Organization of Sounds* (MIT Press, Cambridge, MA, 1990).
- Feng, A. S. & Ratnam, R. *Annu. Rev. Psychol.* **51**, 699–725 (2000).

Trans-complex formation by proteolipid channels in the terminal phase of membrane fusion

Christopher Peters^{*†}, Martin J. Bayer^{*†}, Susanne Bühler^{*}, Jens S. Andersen[‡], Matthias Mann[‡] & Andreas Mayer^{*}

^{*} Friedrich-Miescher-Laboratorium der Max-Planck-Gesellschaft, Spemannstrasse 37-39, 72076 Tübingen, Germany

[‡] Department of Molecular Biology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

[†] These authors contributed equally to this work

SNAREs (soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors) and Rab-GTPases, together with their cofactors, mediate the attachment step in the membrane fusion of vesicles. But how bilayer mixing—the subsequent core process of fusion—is catalysed remains unclear. Ca^{2+} /calmodulin controls this terminal process in many intracellular fusion events. Here we identify V0, the membrane-integral sector of the vacuolar H^+ -ATPase, as a target of calmodulin on yeast vacuoles. Between docking and bilayer fusion, V0 sectors from opposing membranes form complexes. V0 *trans*-complex formation occurs downstream from *trans*-SNARE pairing, and depends on both the Rab-GTPase Ypt7 and calmodulin. The maintenance of existing complexes and completion of fusion are independent of *trans*-SNARE pairs. Reconstituted proteolipids form sealed channels, which can expand to form aqueous pores in a Ca^{2+} /calmodulin-dependent fashion. V0 *trans*-complexes may therefore form a continuous, proteolipid-lined channel at the fusion site. We propose that radial expansion of such a protein pore may be a mechanism for intracellular membrane fusion.

Rab-like GTPases and associated tethering factors and SNARE proteins mediate membrane attachment in intracellular fusion reactions^{1,2}. After their activation (priming) by Sec18/NSF (*N*-ethylmaleimide-sensitive factor) and Sec17/SNAP (soluble NSF attachment protein), SNAREs can form complexes in *trans*, that is, between adjacent membranes, which probably leads to tight membrane apposition. Two main models have been proposed for subsequent bilayer fusion^{1,3–5}. ‘Proximity’ models suggest that forcing the membranes into intimate contact suffices to mediate rapid spontaneous mixing of the bilayers, probably through hemifusion intermediates. ‘Fusion pore’ models assume a continuous proteinaceous channel connecting the two membranes that forms a central hydrophilic pore. Such a channel could mediate fusion if it were composed of several subunits, and if the subunits could separate and move apart radially. This would enlarge the hydrophilic lumen and create an amphipathic surface between the subunits. Lipids could invade these amphipathic spaces enabling them to flow between the membranes.

A proximity model has been proposed for endoplasmic fusion events⁶. In SNARE complexes the carboxy-terminal transmembrane regions of SNAREs are in close proximity⁶, as in viral fusion proteins⁷. It was suggested that SNARE complex formation, or the rearrangement of viral fusion proteins, helps to closely appose and fuse the membranes⁶. SNAREs mediate fusion of synthetic liposomes⁸, but other purified proteins, such as p97/cdc48 or NSF, produce similar effects^{9,10}. In most physiological membrane systems, for example exocytosis, *trans*-SNARE pairs have not been assayed, which has led to conflicting conclusions on their role. Exocytosis in sea urchin eggs proceeds in the absence of detectable SNARE complexes¹¹, whereas studies in PC12 or chromaffin cells have indicated that different conformations or association states of SNAREs exist in the primed state^{12,13}. If interpreted in the light of *trans*-SNARE pairing, these results are consistent with a conformational rearrangement of *trans*-SNARE complexes driving bilayer mixing. However, this interpretation is less straightforward if the associations of SNAREs with many other proteins, which also participate in fusion^{1,2}, are considered. This leaves room for models

in which *trans*-SNARE complexes do not have a direct role in bilayer fusion.

Trans-SNARE complexes were assayed in vacuole fusion. There they can be disassembled before full fusion without preventing completion of the reaction^{14,15}. Completion depends on calmodulin, an efflux of luminal Ca^{2+} (ref. 16) and protein phosphatase 1 (ref. 17). Calmodulin and an efflux of luminal Ca^{2+} also control membrane fusion at other stations of intracellular trafficking, underscoring their general relevance^{18–21}. Here we have isolated the vacuolar binding partners of calmodulin and characterized their interactions.

Calmodulin binding proteins on vacuolar membranes

Ca^{2+} enhances calmodulin binding to vacuole membrane¹⁶. We sought to identify proteins in the vicinity of calmodulin by label transfer experiments in combination with mass spectrometric protein analysis. Purified calmodulin was covalently coupled to iodinated SASD (see Methods), a cleavable, light-activatable, hetero-bifunctional crosslinker. This modified calmodulin was added to vacuoles, allowing it to bind to the membranes. At different times of fusion, the vacuoles were irradiated with ultraviolet to induce crosslinks to neighbouring proteins. The membranes were re-isolated and solubilized. The crosslinker was cleaved by reduction to release the crosslinked proteins with only a small covalent modification carrying the iodine label.

Nine bands could be distinguished reproducibly on autoradiograms of SDS gels (Fig. 1a). One of the bands was calmodulin, which had transferred label to itself—a reaction that could also be seen in the absence of membranes (data not shown). The other bands increased in intensity over the course of the reaction. Two proteins of relative molecular mass 50,000 to 60,000 (M_r 50–60K) were labelled already in the initial phase of the reaction and showed a smaller increase from that level than the others. The crosslink products depended on fusion. They were not generated in the presence of various inhibitors of the reaction (Fig. 1b): Apyrase, which depletes the ATP regenerating system; BAPTA, a chelator of calcium that is released from the vacuolar lumen in response to the

docking event; W7, a calmodulin inhibitor¹⁶; Gdi1, an inhibitor of docking that acts on the Rab-like GTPase Ypt7 (refs 14, 22); or antibodies to Sec18, which prevent SNARE activation^{22,23}.

The label transfer experiments were performed on a preparative scale and the vacuolar proteins were separated by two-dimensional polyacrylamide gel electrophoresis (2D PAGE) using the cationic detergent 16-BAC in the first and SDS in the second dimension²⁴. Labelled protein spots were excised from the gel, the proteins were degraded by trypsin, and the resulting peptide mixture was analysed by matrix-assisted laser-deionization mass spectrometry (MALDI-MS) and nanoelectrospray MS/MS as described¹⁷ (Fig. 2).

For some spots (n.d.) the label was not on the major protein in that area. The other spots yielded several hits on different components of the vacuolar H⁺-ATPase (V-ATPase), namely Vph1, Vma6 and Vma2. Three spots identified Vtc4 (or fragments thereof, respectively), a protein that has been implicated in V-ATPase function²⁵. Three spots yielded spectra for Gpd1 (glyceraldehyde-3-phosphate-dehydrogenase) or Eno2 (enolase), i.e. glycolytic enzymes. A functional relationship of these proteins to the V-ATPase is possible. The V-ATPase consists of a peripheral V1

sector which is composed of at least eight subunits, including Vma2, and carries the ATPase activity. V1 is bound to the membrane integral V0 sector containing Vma6, Vph1, and the proteolipids Vma3 (six copies), Vma11 and Vma16 (ref. 26). The V1 sector can be shed from the V0 sector and become soluble. Disassembly is triggered under conditions of low glucose by a mechanism dependent on glycolysis downstream of glucose-6-phosphate²⁷. Gpd1 and/or Eno2 might be involved in disassembly.

The 16-BAC/SDS system could not resolve the crosslink product of ~17K that is visible in Fig. 1. Membrane lipids interfere with the 16-BAC gel and must be extracted with chloroform/methanol. In this step the 17K band was lost into the organic extract, that is, it behaved like a proteolipid. Proteolipids readily partition into organic solvents like chloroform/methanol, in which virtually all other proteins are insoluble. The organic extract of the vacuoles consisted essentially only of protein migrating at 17K, as shown by Coomassie staining (Fig. 3a). This protein was not present if vacuoles from a mutant lacking the proteolipid gene Vma3 were extracted (data not shown). The extracted protein was radioactive and comigrated with the crosslink product of 17K present on autoradiographs of whole vacuolar membranes. Therefore, the 17K crosslink product represents V-ATPase proteolipid.

A co-adsorption experiment confirmed this. Vacuolar membranes were solubilized in Triton X-100 and fractionated on a column containing immobilized calmodulin antibodies. After extensive washing, so that only the tightest and most direct interactors of calmodulin were retained on the resin, specific proteins in the 17K range could be eluted (Fig. 3b). These bands were absent if the vacuoles had been preincubated with the calmodulin inhibitor W7. In a preparative-scale experiment, the 17K band was identified by nanoelectrospray MS/MS as a mixture of Vma3 and Nr1/Vtc1 (Fig. 3c). Like Vtc4, Nr1 is involved in V-ATPase function, but by an unknown mechanism²⁵. Purified proteolipid, prepared by organic extraction of vacuoles, could also bind calmodulin when coated onto enzyme-linked immunoassorbent assay (ELISA) plates (Fig. 3d). Binding was weakened by the calcium chelator EGTA and

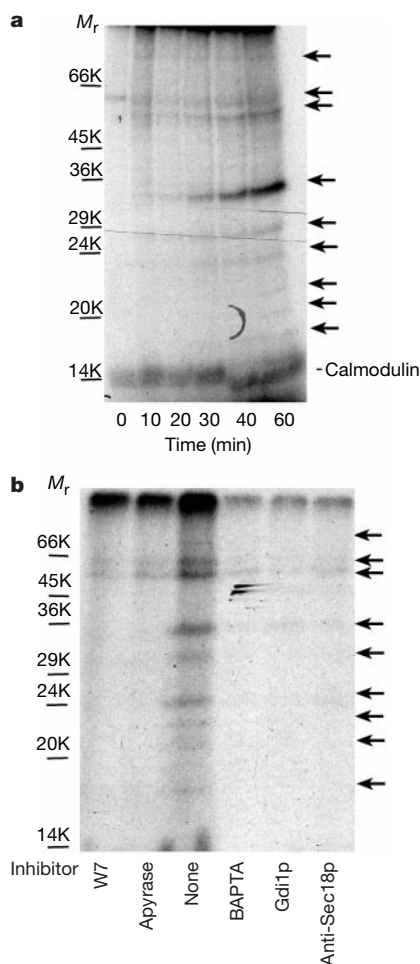


Figure 1 Label transfer from iodinated calmodulin to vacuolar membrane proteins. Vacuoles from BJ3505 (100 µg) were incubated in fusion reactions with 0.8 µM calmodulin labelled with iodinated SASD. **a**, At the indicated times, vacuoles were re-isolated (2 min, 10,000g, 2 °C), resuspended in 100 µl of PSK buffer containing 100 µM CaCl₂ and 0.5 mM MnCl₂, and irradiated with ultraviolet. The vacuoles were diluted to 1 ml with the same buffer, re-isolated as before and analysed by SDS-PAGE and autoradiography. **b**, As in **a**, except that the fusion samples were incubated for 45 min in the presence of BAPTA (3 mM), Gdi1 (2.5 µM), affinity-purified antibodies to Sec18 (1.5 µM), apyrase (–ATP; 20 U ml^{–1}; Sigma, grade VIII), or W7 (150 µM). Arrows indicate reproducibly labelled bands.

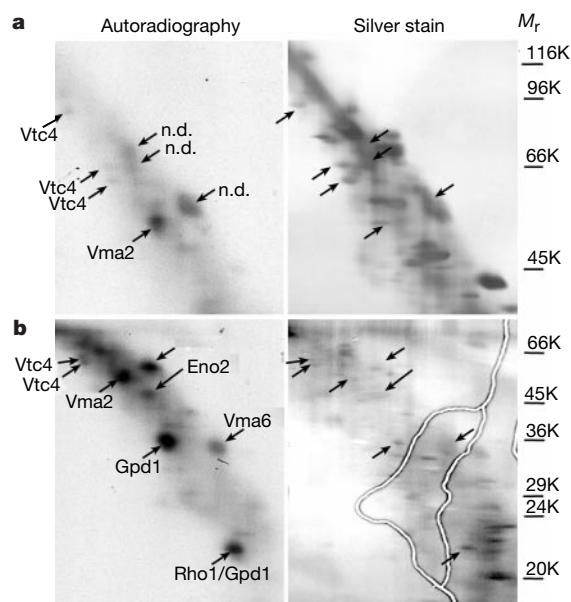


Figure 2 Two-dimensional separation of proteins labelled by SASD-calmodulin. Vacuoles from BJ3505 (1 mg) were incubated (45 min) in a fusion reaction with SASD-calmodulin, re-isolated and irradiated as in Fig. 1. The final pellet was analysed by 2D 16-BAC/SDS-PAGE on 8% (**a**) or 14% (**b**) gels, silver staining and autoradiography. Well-separated labelled spots were excised and analysed by MALDI-MS. In some spots (n.d.) the labelled protein was overlaid with other polypeptides, preventing identification of the labelled species.

did not occur when only BSA was coated onto the surface as a control protein. Because Nrfl does not partition into the organic extract (M.B., unpublished data), the relevant calmodulin-binding partner in the extract appears to be Vma3.

Calmodulin-dependent pore formation by proteolipids

Proteolipids of the V0 sector form proton tight channels composed of at least six subunits^{26,27}. V-ATPase proteolipid channels from Torpedo electric organ are capable of forming pores that can release acetylcholine, in other words, a molecule considerably larger than H⁺, in a Ca²⁺-dependent fashion²⁸. Because this preparation contained at least one contaminant of 14K (ref. 29), it is uncertain whether the

contaminant or the proteolipid itself mediates the Ca²⁺ effect. We purified proteolipids from yeast vacuoles by immunoprecipitating the V-ATPase and extracting the proteolipids from the immunoprecipitate with chloroform/methanol. The proteolipids were reconstituted into liposomes loaded with choline. Channel opening was monitored through choline release in a coupled chemiluminescent reaction with choline oxidase, peroxidase and luminol.

Choline was rapidly released on addition of Ca²⁺ and calmodulin to these vesicles (Fig. 3e). Release did not occur if either calmodulin or Ca²⁺ was omitted, or if Ca²⁺ was replaced by Mg²⁺. Liposomes without proteolipids remained tightly sealed under all conditions (data not shown). This resembled the properties of reconstituted

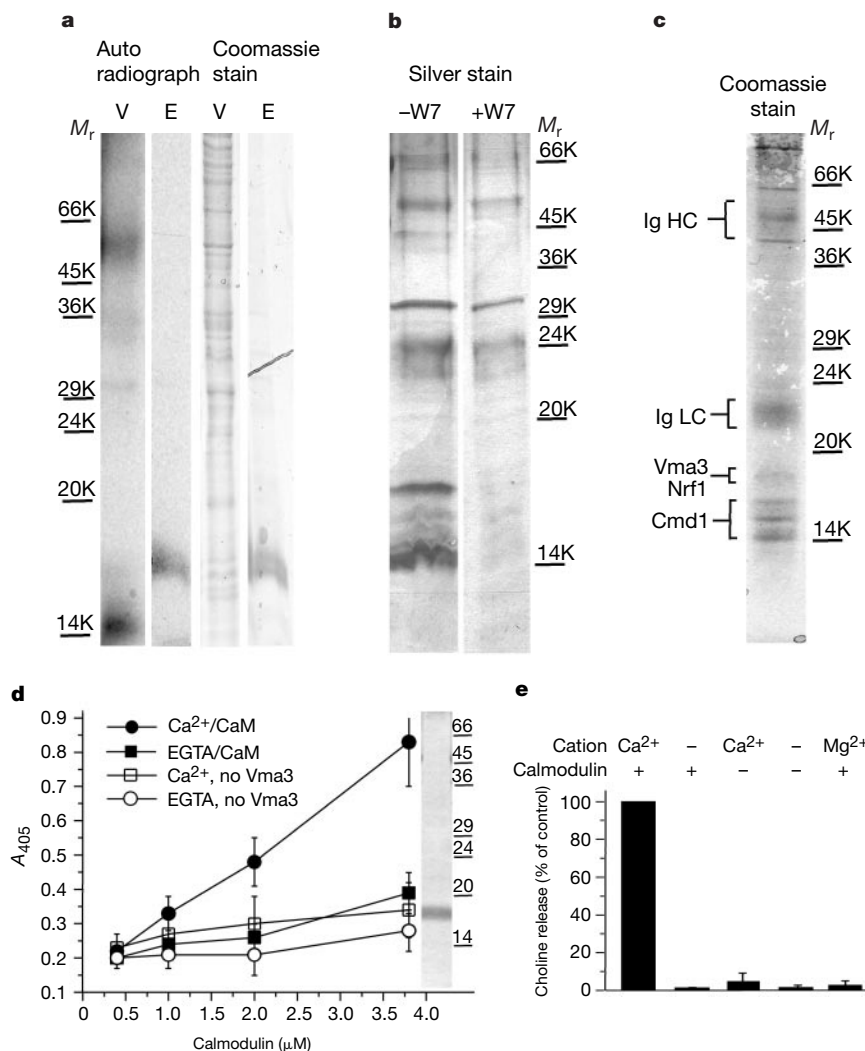


Figure 3 Calmodulin–proteolipid interaction. **a**, BJ3505 vacuoles (5 mg) were used in a label transfer experiment as in Fig. 2. After extraction with chloroform/methanol, proteins from the organic phase (E) were separated by SDS–PAGE. Vacuolar membranes (V; 50 μg) are included for comparison. **b**, Co-immunoprecipitation on anti-calmodulin beads. BJ3505 vacuoles (5 mg) were incubated in 10 mM PIPES/KOH pH 6.8, 1 mM PMSF (10 min, 0 °C), re-isolated (4 min, 20,000g, 2 °C), resuspended in 5 ml of PSK, 0.5 mM MnCl₂ and divided into two. Membranes were washed twice in PSK with 1 mM W7 (or buffer only), 0.5 mM MnCl₂ (5 min, 27 °C), re-isolated and solubilized in 5 ml of 1% Triton X-100 in PSK, 100 μM CaCl₂, 0.5 mM MnCl₂. The lysate was clarified (5 min, 100,000g) and shaken with 100 μl of anti-calmodulin Sepharose (1 h, 4 °C). Beads were column washed with 25 ml of solubilization buffer (with 0.5% Triton X-100) and prepared for SDS–PAGE. **c**, Preparative co-immunoprecipitation on anti-calmodulin beads. Vacuoles (20 mg) were extracted (except for the W7 washes), solubilized, immunoprecipitated and analysed as in **b**. Bands from the 14–20K region were excised and identified by MALDI-MS. Note that proteolipids stain very poorly with Coomassie. **d**, Ca²⁺-dependent

calmodulin–Vma3 interaction. The chloroform/methanol extract from 5 mg of vacuoles was concentrated to 50 μl by evaporation, diluted with 0.5 ml of buffer S (1% Triton X-100 in PSK, 100 μM CaCl₂) and adsorbed to ELISA plates (15 h, 4 °C). Controls were buffer S only. After blocking with 2% BSA (1 h, 4 °C), vials were incubated with the following solutions in buffer W (buffer S, 2% BSA, 0.5% Triton X-100): different concentrations of calmodulin (2 h, 23 °C); buffer (3 washes); anti-calmodulin IgG (1 h, 23 °C); alkaline-phosphatase-coupled goat-anti-rabbit IgG (1 h, 23 °C); buffer W with or without 1 mM EGTA (3 washes); alkaline phosphatase assay buffer (as for the fusion assay). Alkaline phosphatase activity was quantified by the A₄₀₅ of *p*-nitrophenol (*n* = 3). **e**, Choline release. Proteolipid-containing liposomes were prepared and loaded with choline. Where indicated, 100 ng ml⁻¹ calmodulin had been reconstituted together with the proteolipids. Ca²⁺, Mg²⁺ or only control buffer were injected. The initial rate of release was quantified as change of luminescence. Less than 10% of enclosed choline was released in the first 60 s after injection. Ca²⁺/calmodulin was the 100% reference (*n* = 3).

proteolipids from neuronal tissue which released acetylcholine in response to Ca^{2+} (ref. 30). We conclude that calmodulin interacts with V0-proteolipid channels and can induce large conformational changes on them. As each proteolipid subunit consists of only four transmembrane helices connected by very short loops, pore opening will be most likely to involve radial expansion of a central hydrophilic cavity between the subunits. Thus the proteolipids possess an important property postulated in a pore model of membrane fusion⁴.

A V-ATPase function beyond that as a proton pump

Vacuole fusion requires an electrochemical membrane potential that is mainly generated by the V-ATPase^{15,33}. In line with that, vacuoles from V-ATPase mutants could not fuse (data not shown). The electrochemical potential is required up to the docking stage^{15,22,31}, probably for *trans*-SNARE pairing¹⁵. If Vma3 is the target of calmodulin, which acts in the post-docking phase¹⁶, we must postulate a later function for at least the V0 sector that is independent of proton translocation.

We developed conditions for isolating vacuoles with an electrochemical potential, but devoid of H^+ -pump activity. Vacuoles were briefly incubated with ATP to generate a H^+ gradient and then extracted with SCN^- (refs 26, 27). This completely removed V1 sectors (Fig. 4a, inset). Such vacuoles maintained the H^+ gradient because V0 sectors devoid of the V1 heads do not conduct protons²⁷. The H^+ -pump activity of SCN^- -extracted vacuoles was reduced to background levels, that is, it became indistinguishable from that measured without ATP, or in the presence of the protonophore carbonyl cyanide-4-trifluoromethoxyphenylhydrazone (FCCP) (Fig. 4a). However, the vacuoles were still fusion competent (Fig. 4b). Fusion occurred along the authentic pathway because it was sensitive to Microcystin LR, a late-acting inhibitor of the reaction³¹ (Fig. 4b). In agreement with their significant but reduced fusion activity (Fig. 4b), the pre-established gradient of SCN^- -extracted vacuoles was lower than that of non-extracted control vacuoles (Fig. 4c), which could actively pump protons (compare Fig. 4a). Quinacrine accumulation and fusion competence of the SCN^- -extracted vacuoles were abolished by the uncoupler FCCP, confirming that the pre-existing H^+ gradient was required for fusion (Fig. 4b, c).

The V-ATPase inhibitor DCCD inactivates the V0 sector by binding to Glu 137 of Vma3, which is most probably situated on the periphery of the hexameric Vma3 channel³². As with SCN^- extraction, DCCD abolished the H^+ -pump activity (Fig. 4a), but did not dissipate the existing H^+ gradient (Fig. 4c). In contrast to SCN^- extraction, however, DCCD inactivated the vacuoles for fusion (Fig. 4b). We conclude that vacuoles remain fusion competent in the absence of H^+ -pump activity during the reaction. The V0 sector must therefore have a function for fusion that is independent of H^+ -translocation.

V0-V0 *trans*-complexes form after docking

We used tagged V0 subunits to investigate interactions between V0 sectors from docked vacuoles by solubilization and rapid co-immunoprecipitation. One fusion partner contained the V0 subunit Vph1 with a hexahistidinyl haemagglutinin A (His_6 -HA₃) tag; the other fusion partner contained the V0 subunit Vma3 with an AU1 tag. We detected an association of Vma3-AU1 with Vph1- His_6 -HA₃ which increased during the reaction, suggesting the formation of V0 complexes (Fig. 5a). Combinations of Vph1-AU1 and Vph1- His_6 -HA₃ (Fig. 5c) or Vma3-AU1 and Vma3-HA (data not shown) yielded similar results. We routinely used the Vph1-AU1/Vph1- His_6 -HA₃ combination, in which the tags were on the cytosolic face. The tags on Vma3 were exposed to the lumen of the vacuole and were degraded to significant degrees, requiring large amounts of vacuoles for detection.

When V0 sectors with two different tags were co-expressed on

the same vacuole they co-precipitated only if such vacuoles fused, and not if fusion was blocked by the docking inhibitor Gdi1, nor if freshly isolated vacuoles were used (Fig. 5b). Therefore, V0 complexes occurred only during the fusion reaction. They were

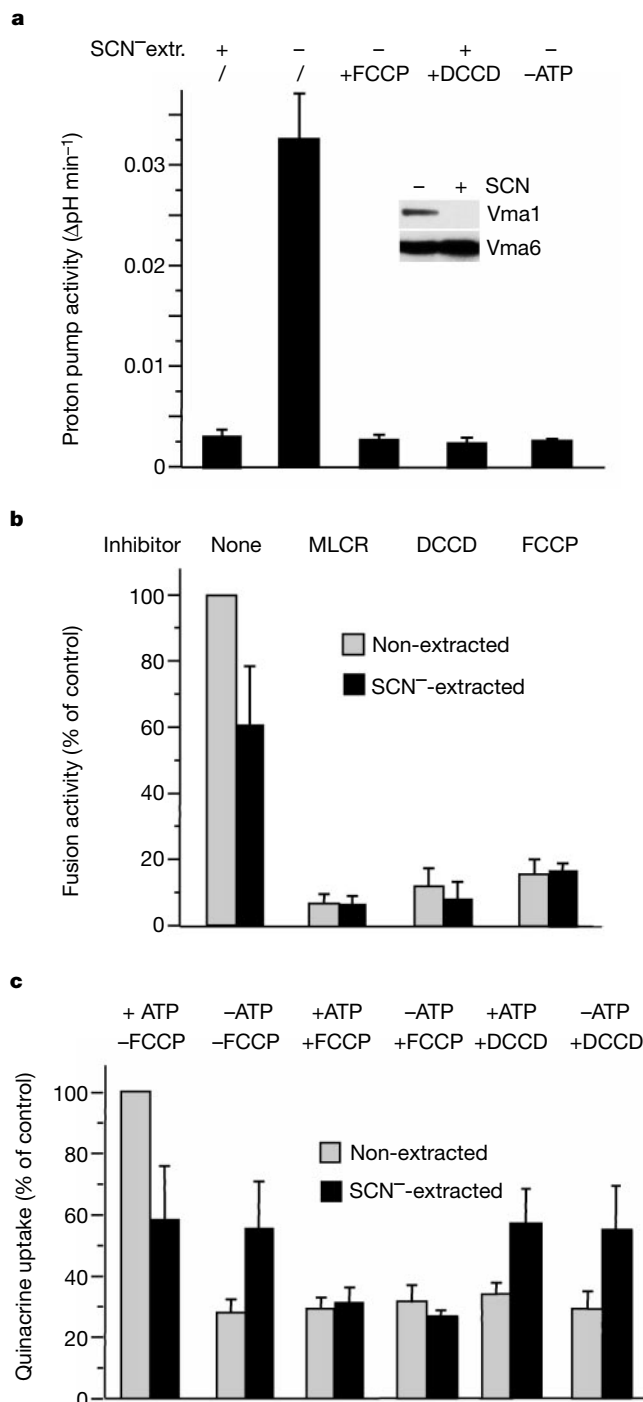


Figure 4 V-ATPase pump activity is not directly required for vacuole fusion. **a**, Vacuoles were extracted with SCN^- or left untreated. Their H^+ -pump activity was determined in the absence or presence of FCCP (10 μM), DCCD (50 μM), or the ATP regenerating system ($n = 6$). Inset, an aliquot of the vacuoles from one of the experiments was assayed by immunoblotting for the presence of the V1 subunit Vma1 and the V0 subunit Vma6. **b**, Fusion activity of SCN^- -extracted and untreated vacuoles was determined in the presence or absence of FCCP (30 μM), DCCD (50 μM), or Microcystin LR (10 μM) ($n = 6$). The non-extracted sample without inhibitors was the 100% reference. Fusion activities for these samples were 2–4 U. **c**, The proton gradient across SCN^- -extracted and untreated vacuoles was assayed by quinacrine uptake under the same conditions as in **a** ($n = 4$). The samples with ATP and no FCCP served as 100% reference; their A_{430} was 0.3–0.5.

not formed by association of differentially tagged V0 complexes *in cis*, that is, on the same membrane, which might occur, for example, after fusion.

Most inhibitors of fusion prevented the formation of V0 *trans*-complexes (Fig. 5c): for example, the docking inhibitor Gdi1 (ref. 22); the uncoupler FCCP¹⁵; the Ca²⁺ chelator BAPTA, which quenches an efflux of luminal Ca²⁺ from the vacuoles which is triggered by docking¹⁶; or the calmodulin inhibitor W7 (ref. 16). W7 inhibited complex formation more thoroughly than BAPTA (Fig. 5c) suggesting that the initial function of calmodulin in the formation of V0 complexes (blocked by W7) may be separable from a subsequent Ca²⁺-induced conformational change (blocked by BAPTA) on the *trans*-complexes. Further detailed analysis will be required to resolve this issue. Complexes of V0 did accumulate in the presence of two very late acting inhibitors, the protein phosphatase 1 (PP1) inhibitor Microcystin LR^{17,31} (Fig. 5c) and GTP-γS (Fig. 5d).

To specifically monitor the late effect of GTP-γS, and not its early effects on docking (A.M., unpublished data), we accumulated a fusion reaction in the post-docking stage by adding BAPTA. When the vacuoles were released from this block fusion was still sensitive to GTP-γS and Microcystin LR, but V0 complex formation was not affected. Our observations allow two firm conclusions: first, V0 complexes originated from V0 sectors in opposing membranes in the absence of fusion; second, this *trans*-complex formation kinetically maps to a very late step of fusion. It requires Ypt7 (docking) and Ca²⁺/calmodulin, which acts in the transition from docking to bilayer fusion¹⁶. However, *trans*-complex formation precedes the PP1-dependent and the late GTP-γS-sensitive step.

Which proteins are associated with V0 *trans*-complexes? We fused vacuoles carrying Vph1–AU1 with vacuoles carrying Vph1–His₆–HA₃ in the presence of Microcystin to block fusion. The membranes were solubilized and V0 sectors were isolated by adsorption to Ni²⁺–NTA resin. Ni²⁺–NTA bound to V0 monomers containing only the Vph1–His₆–HA₃ from one fusion partner as

well as the *trans*-complexes which contained also Vph1–AU. Bound V0 complexes were eluted with imidazole and immunoadsorbed to anti-AU1 Sepharose (selecting for *trans*-complexes) or anti-HA Sepharose (re-isolating predominantly monomers). Comparison of the relative amounts of coadsorbed proteins from the monomers and the *trans*-complexes indicated that calmodulin and the Q-SNARE Vam3 were strongly enriched on the *trans*-complexes (Fig. 5e). The V1 sector of the V-ATPase was present in the monomer fraction but absent from the *trans*-complexes.

SNARE pairing precedes V0–V0 *trans*-complexes

In comparison to the t-SNARE Vam3, we routinely found much smaller fractions of the R-SNARE Nyv1 associated with V0 *trans*-complexes (Fig. 6a), and levels were often barely distinguishable from nonspecific background. When we could detect a signal, however, this co-precipitation could be abolished by incubating the vacuoles with Sec18 (NSF) and Sec17 (α-SNAP). An excess of Sec18/NSF and Sec17/α-SNAP disassembles *trans*-SNARE pairs^{14,15}. This prevents fusion if carried out before docking, but not after docking is completed, in other words, at a stage when *trans*-SNARE pairs have been present for some time. The late, Ca²⁺- and PP1-dependent stage of fusion^{16,17} proceeds independently of *trans*-SNARE pairing suggesting an intermediary function for *trans*-SNARE pairs. If added at the beginning of a fusion reaction, excess Sec18/Sec17 suppressed the formation of V0 complexes and fusion (Fig. 6b). After 40 min (completion of docking²²), both V0 complexes and fusion were resistant to Sec18/Sec17. The post-docking phase inhibitor BAPTA permits *trans*-SNARE pairing¹⁵, but prevents formation of V0 *trans*-complexes. In the presence of BAPTA the vacuoles reached a state that, after removal of BAPTA, allowed V0 *trans*-complex formation, even if excess Sec18/Sec17 was added to this second incubation (Fig. 6c). V0 *trans*-complexes also resisted disassembly by Sec18/Sec17 if these proteins were added into the detergent extract together with Mg-ATP (Fig. 6c), which excludes the possibility that the resistance

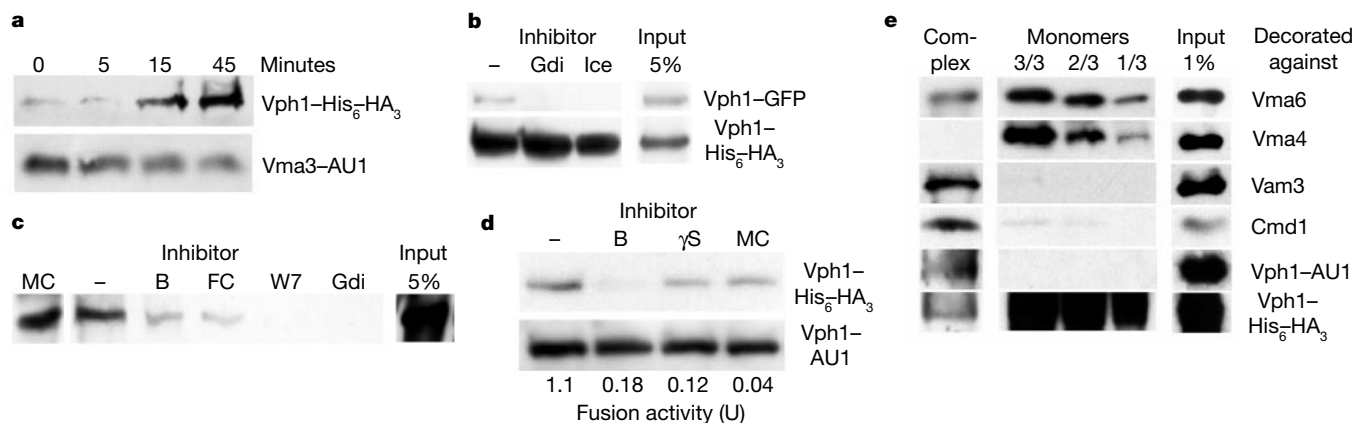


Figure 5 Complex formation of V0 sectors. **a**, Time course. BJ3505 vacuoles bearing Vma3–AU1 or Vph1–His₆–HA₃ were incubated at 27 °C in 1-ml fusion reactions. Vacuoles were then re-isolated and solubilized in buffer I. Vma3–AU1 was immunoprecipitated using monoclonal mouse anti-AU1. Co-immunoprecipitated Vph1–His₆–HA₃ was analysed by western blotting with rat anti-HA. **b**, Vph1–His₆–HA₃ and Vph1–GFP were co-expressed in BJ3505 cells. Vacuoles from these cells were fused (70 min, 27 °C) in the presence or absence of 2.5 μM Gdi1 or left on ice, and V0 complex formation was analysed as in **a**. **c**, Fusion reactions with BJ3505 vacuoles bearing Vph1–AU1 or Vph1–His₆–HA₃ were incubated (90 min) in the presence of 10 μM Microcystin LR (MC), 2.5 μM Gdi1, 150 μM W7, 30 μM FCCP (FC), 5 mM BAPTA (B), or buffer. *Trans*-complex formation was analysed as in **a**. **d**, A fusion reaction as in **c** was incubated (30 min, 27 °C) in the presence of 5 mM BAPTA. The vacuoles were re-isolated, resuspended in fusion buffer with 150 μM CaCl₂ and supplemented with 5 mM BAPTA (B), 10 μM Microcystin LR, or 2 mM GTP-γS (γS). After a second incubation, fusion and

V0 complex formation were assayed as in **a**. **e**, Characterization of *trans*-complexes. A 10-ml fusion reaction was carried out as in **c**, except with 10 μM Microcystin LR. Vacuoles were re-isolated, solubilized in buffer D (buffer I with 5% glycerol, 0.1 μM calmodulin, 0.6 mg ml⁻¹ cytosol, pH 7.5) and centrifuged (5 min, 100,000g). The supernatant was incubated with 125 μl Ni-NTA agarose (1 h, 4 °C), apart from 1% which was kept to use as an input standard. After three washes with buffer D, proteins were eluted with 2 ml of 70 mM imidazole in buffer D (with only 100 mM KCl), and 1.5 ml was re-precipitated (45 min, 4 °C) with 30 μg of anti-AU1 (*trans*-complexes) and 0.5 ml was re-precipitated with 60 μg of mouse monoclonal anti-HA (monomers). The beads were washed three times with 0.5% Triton X-100, 100 μM CaCl₂ in PSK, suspended in SDS buffer and analysed by western blot. The HA eluate (monomers) was applied in three different amounts (30, 20 and 10 μl) to facilitate comparison with the AU1 eluate (complex).

is due to limited accessibility of *trans*-SNARE pairs to Sec18/Sec17 (ref. 33). In summary, *trans*-SNARE pairing was required to form *trans*-V0 complexes, but not to maintain them.

Discussion

Inactivation of V-ATPase genes impairs intracellular trafficking, along both the biosynthetic and the endocytic routes^{34–37}. V0 is

present in all compartments of the biosynthetic and endocytic pathways. It is assembled in the endoplasmic reticulum and then distributed over the Golgi, endosomes, lysosomes and transport vesicles. It can reversibly dissociate from the V1 sector²⁷. Deletion of the V0 proteolipid Vma3 inhibits trafficking along the biosynthetic pathway more profoundly than do mutations in V1 (refs 36, 37), although both abolish proton pump activity²⁶. This is consistent with our findings that suggest that V0, but not V1, has a direct role independent of proton pumping in the fusion reaction. Furthermore, calmodulin and an efflux of luminal calcium, crucial factors for V0 *trans*-complex formation and vacuole fusion¹⁶, are also required in endosome fusion, endosome–lysosome fusion and intra-Golgi transport^{18–21}.

Calmodulin enhances the triggering phase of exocytosis^{38,39} and SNAREs in synaptosome extracts were found in association with V0 and with synaptophysin^{40,41}. This could involve V0 subunits from both membranes because V-ATPases were detected in plasma membranes^{29,42} and every fusion of secretory vesicles transfers V0 subunits from the vesicle membrane into the plasmalemma. On the basis of these observations V0 subunits might operate in various trafficking steps. But it is equally possible that distinct proteins with properties similar to those of vacuolar proteolipids perform related functions on other compartments. For example, the Golgi SNARE Sed5 is genetically linked to Got1 and Sft2 (ref. 43), which act in ER–Golgi and endosome–Golgi trafficking. Got1 functions after membrane attachment. Like the V0 proteolipids, Got1 and Sft2 are very small and hydrophobic, have four or five transmembrane domains and possibly form multimers⁴³.

V0 mutants can grow slowly on acidic, rich media (although they die on neutral media²⁶). In contrast, mutations in, for example, SNAREs or Rabs can be lethal². A possible involvement of V0 in transport would have to be reconciled with the non-lethality of V0 mutations. Several assumptions can be made. Functional redundancy is possible, enabling yeast cells to compensate for the loss of V-ATPase subunits. In support of this, a temperature-sensitive mutant of the V1 subunit Vma4 showed more severe phenotypes than the Vma4 deletion mutant, although both mutations abolished proton pumping²⁷. Furthermore, suppression of V-ATPase mutants does occur²⁵. Suppressors of V0 mutations are much rarer than those for V1 mutations²⁵, although both V0 and V1 mutations completely eliminate H⁺-pump activity^{26,27}. This supports our observation of an additional function of V0 that is independent of proton translocation.

Our preferred interpretation is kinetic. Inactivation of the SNARE or Rab systems interferes with membrane attachment. Membrane contacts should then be reduced to short accidental collisions providing insufficient time, stability and proximity for productive fusion. If the attachment systems work and downstream events are slowed down by inactivation of their enzymes, then the membranes should remain docked. Slow spontaneous bilayer fusion may ensue after some time, providing a reduced reaction rate that may nevertheless suffice to promote slow growth. Such mutations should become lethal under conditions impeding spontaneous fusion, such as low temperature. Cold sensitivity is indeed a feature of Vma mutants²⁶.

Our results can be integrated into a pore model^{14,44} of membrane bilayer mixing. *Trans*-SNARE pairing (docking) and calmodulin serve to assemble V0 sectors from apposed vacuolar membranes into *trans*-complexes, which may create a proteolipid-lined channel spanning both membranes. *Trans*-SNARE pairs are dispensable from this point, although they may persist and still help to stabilize the tight apposition of the membranes. Efflux of luminal Ca²⁺, activated by docking, may then trigger a conformational change of proteolipids and open a hydrophilic pore. Proteolipids are small. Pore opening should therefore involve a radial expansion of the oligomeric proteolipid ring by lateral separation of its subunits, which creates amphiphilic clefts. Hydrophilic lipid headgroups

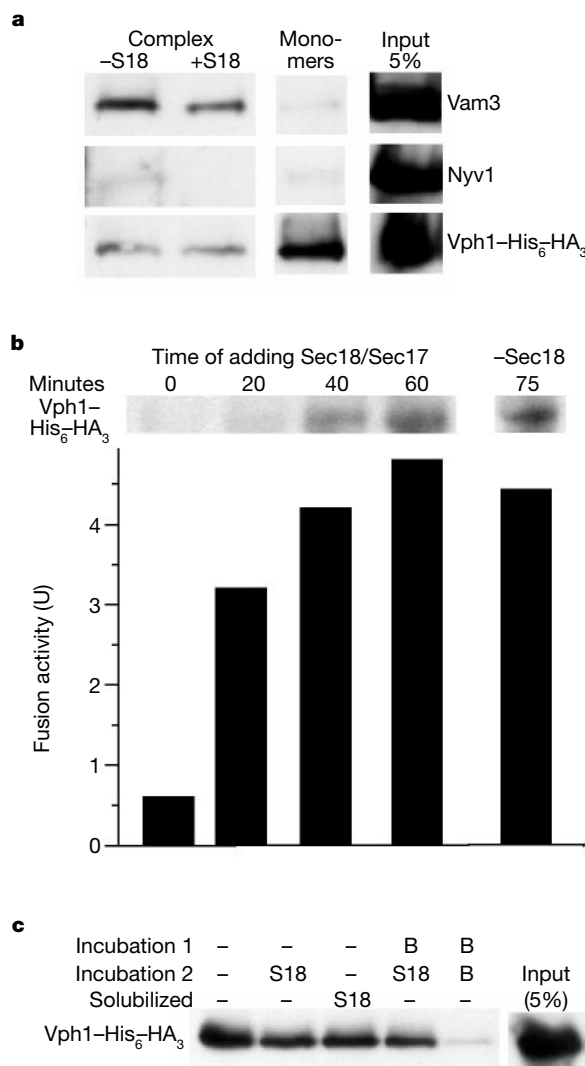


Figure 6 *Trans*-SNARE pairing and V0 *trans*-complexes. **a**, BJ3505 vacuoles bearing Vph1-AU1 or Vph1-His₆-HA₃ were mixed and fused in 10-ml reactions (60 min) with 10- μ M Microcystin LR. Sec18/NSF (60 μ g ml⁻¹) and Sec17/SNAP (6 μ g ml⁻¹) (+S18) or buffer (-S18) were added for 15 min at 27 °C. The vacuoles were re-isolated, solubilized and analysed for V0 *trans*-complexes and monomers as in Fig. 5d. Monomers were isolated only from a sample without Sec18/Sec17. **b**, Time course of Sec18/Sec17 addition. Fusion reactions (1 ml each) were carried out as in **a** for 75 min, and 60 μ g ml⁻¹ Sec18/NSF and 6 μ g ml⁻¹ Sec17/SNAP were added as indicated. The vacuoles were re-isolated, solubilized and assayed for V0 *trans*-complexes as in Fig. 5b. In parallel, small-scale standard fusion reactions were run without Microcystin, but with the same amounts of Sec18/Sec17. Fusion was assayed after 75 min. A twofold concentrated ATP-regenerating system was used. **c**, Fusion reactions (1 ml) as in **a** were performed in two steps with the indicated additions of B (5 mM BAPTA), S18 (120 μ g ml⁻¹ Sec18, 12 μ g ml⁻¹ Sec17), or only buffer (-). After the first incubation (30 min, 27 °C), the vacuoles were re-isolated, resuspended in fusion buffer and incubated again. Samples without BAPTA received 50 μ M CaCl₂. Vacuoles were re-isolated, solubilized in buffer I with the same ATP-regenerating system as in the fusion buffer, incubated for 10 min at 27 °C, and chilled and analysed for V0 *trans*-complexes as in Fig. 5b. All samples contained Microcystin LR (10 μ M).

could migrate into the lumen of the pore while the hydrophobic fatty-acid chains could fill the space between the transmembrane helices. Dissociation of the *trans*-complexes into subunits would allow irreversible expansion of the fusion pore and reset the system. Pore expansion and/or dissociation of the *trans*-complexes into monomeric subunits might be controlled by protein phosphatase 1.

The hypothesis of radial opening of a protein pore (ROPP) can accommodate experimental observations from different systems. First, it ascribes a molecular function to *trans*-SNARE complexes that is in accordance with their time of action in vacuole fusion—a physiological system where *trans*-SNARE pairs could be directly assayed^{14,15}. Kinetic analysis revealed that the SNARE-dependent step probably precedes Ca^{2+} /calmodulin and phosphatase action^{12–17}. *Trans*-SNARE pairs could facilitate the formation of proteolipid *trans*-complexes by forcing the membranes into close contact. They would concentrate V0 sectors at the site of close membrane contact because SNAREs are physically associated with V0 subunits, as shown here in the vacuole system, and in exocytosis^{40,41}. Second, ROPP can explain the flickering of fusion pores by reversible association and dissociation of the pore subunits⁴. Third, a ROPP mechanism could account for the fact that exocytic vesicles can release their contents through a small pore well before full fusion occurs^{45,46}. The dimensions and behaviour of such pores are in the range of large ion channels or gap junctions⁴. Remarkably, a V-ATPase proteolipid hexamer has even been reported as a component of arthropod gap junctions³². Fourth, our results are consistent with observations on exocytosis showing that initial opening of the fusion pore requires calcium but that its expansion is controlled by dephosphorylation⁴⁷. Fifth, ROPP can explain the high speed of specialized exocytotic reactions, such as those that take place in neurons. Pore opening could occur very rapidly by a single Ca^{2+} -triggered conformational change of the pre-existing fusion pore. Sixth, ROPP can explain the large intramembranous particles that are arranged at the preformed exocytosis sites of *Paramecium*. On triggering of fast exocytosis, which is calmodulin- and phosphatase-dependent in this case, they suddenly disperse into six smaller particles⁴⁸. The particles could be radially expanding and disassembling pore complexes. Similarly, a transient expansion of intramembranous particles was observed during exocytosis in *Torpedo* synapses⁴⁹. Last, ROPP could be sensitive to influences of lipid shape on fusion⁵. Lipid flow between the pore subunits might require adoption of a high local membrane curvature. Lipids with a small headgroup and a bulky hydrophobic tail⁵⁰ could be required in such areas on the cytoplasmic leaflet. The luminal leaflet should require a complementary species, lipids with a large headgroup and slim hydrophobic tail⁵.

Trans-SNARE pairing was suggested to force membranes into close contact and drive fusion^{6,8}. Close membrane apposition is probably crucial to physiological fusion and an essential function of SNAREs. In artificial liposome systems, this alone can suffice to induce fusion⁸. The data presented here and earlier results^{11,14–17} indicate, however, that *trans*-SNARE pairing in vacuole fusion is followed by other chemically defined events, such as V0 *trans*-complex formation. To explore the role of V0 *trans*-complexes in the terminal phase of membrane fusion, specific tools need to be developed to interfere with V0 function and monitor changes in its conformation during fusion. □

Methods

Strains and genetic modifications

BJ3505 and DKY6281 are standard fusion strains³¹. His₆–HA₃ and AU1 tags were chromosomally integrated at the 3' end of Vph1 or Vma3 by homologous recombination using PCR products from the following primers and templates. For Vph1–His₆–HA₃ (strain SBY119): forward, 5'-GACATGGAAGTCGCTGTTGCTAGTG-CAAGCTCTTCCGCTTCAAGCT CCCACCAACATCATCATCAC-3'; reverse, 5'-GGTGGATTGGATTGCAAGTCTAAC GTTTTCATGAGATAAGTTTGACTATAGGGA-GACCGGCAGAT C-3'; template pU6H3HA (a gift from D. Gallwitz, Göttingen). For

Vph1-AU1 (strain SBY129): forward, 5'-GACATGG AAGTCGCTGTTGCTAGTG-CAAGCTCTTCCGCTTCAAGCGACACCTATCGCTATATA TAATGAGTCGA-CAACCTTAATATAAC-3'; reverse, as above; template pUG6 (ref. 50). For Vma3–AU1 (strain SBY123): forward, 5'-GCTTGTGTTGTTGAAGTCCAGGCTACTCAAGATGTTGTCTGTGACACCTATCGCTATATAATGAGTCGACACCTTAATATAAC-3'; reverse, 5'-CCCATGTGTAGCTTAAAC ACGCATAATGCTCTTTGTTGCTCCACTATA GGGAGACCGGCAGATC-3'; template pUG6. Gene fusions were verified by western blotting. Plasmid pRS316 expressing Vph1–GFP was provided by R. Piper (Iowa City).

General methods

PSK buffer (10 mM PIPES/KOH pH 6.8, 200 mM sorbitol, 150 mM KCl) was the basis for most solutions. Vacuole isolation and fusion were done as described³¹, except that cytosol was omitted and the vacuole concentration was increased to 0.4 mg ml⁻¹. Polyclonal antisera were raised in rabbits using purified recombinant proteins from *Escherichia coli*. Monoclonal antibodies were purchased for HA (from mouse: Babco; from rat: Roche), AU1 (Babco) and Vph1 (Molecular Probes).

Label transfer

SASD (sulphosuccinimidyl-2-(*p*-azidosalicylamido)ethyl-1,3'-dithiopropionate; Pierce) was made up to 7 mM in DMSO. Aliquots of 10 µl were mixed with 100 µCi ¹²⁵I in 90 µl of PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄, 1.4 mM KH₂PO₄) in an iodogen tube (Pierce) and shaken in the dark (30 s). Calmodulin (900 µl; 55 µmol in PBS) purified from an overexpressing *E. coli* strain¹⁶ was added. After incubation (30 min, room temperature, dark), the buffer was exchanged for PSK buffer by ultrafiltration. The SASD–calmodulin conjugate was frozen in liquid nitrogen at 1 mg ml⁻¹ and stored at –20 °C. SASD–calmodulin was used at 0.8 µM. Crosslinking was induced by ultraviolet from a Zeiss HBO100 Hg-lamp for 3 × 10 s on ice. After re-isolation, the membranes were solubilized in SDS sample buffer with 5% β-mercaptoethanol and boiled for 5 min.

2D PAGE

We extracted vacuolar pellets with chloroform/methanol by the Wessel/Flügge method. Pellets were solubilized in 16-BAC and 1 mg protein per lane was separated on 16-BAC gels (10% acrylamide) as described²⁴. The lanes were equilibrated with SDS stacking gel buffer and loaded on 8% or 14% SDS–PAGE gels for second-dimension analysis²⁴. The gels were silver stained: 1 h in 30% methanol/10% acetic acid, 20 min in 0.4 M sodium acetate pH 6.0, 30% ethanol, 1 mg ml⁻¹ sodium thiosulphate, 3 × 15 min in water, 30 min in 1 mg ml⁻¹ AgNO₃, 0.01% formaldehyde. Development was in 2.5% NaCO₃, 0.015% formaldehyde, stopped by 10 mM EDTA. Stained gels were equilibrated in 5% glycerol, 10% ethanol, dried between cellophane sheets and autoradiographed on Kodak Biomax MS film. Labelled spots were excised and stored in 5% acetic acid.

Mass spectrometry

Proteins were identified as described previously⁵¹. Gel bands were excised and digested in-gel with trypsin. An aliquot of the peptide mixture was analysed by MALDI-MS, using a Bruker Reflex III MALDI-TOF mass spectrometer. For some samples the remaining supernatant was loaded onto Poros R2 and oligoR3 microcartridges and eluted into nanoelectrospray needles. Nano-electrospray MS/MS analysis was performed on a QSTAR quadrupole time-of-flight mass spectrometer (Perkin Elmer). Fragmentation spectra were obtained for as many peptides as possible. Sequence databases were searched with PepSea software (MDS Protana).

Extraction and reconstitution of proteolipids

For label transfer and the plate assay, 5 mg of vacuoles was resuspended in 5 ml 10 mM Tris/HCl pH 7.4, 1 mM EDTA, and centrifuged (5 min, 20,000g, 2 °C). The pellet was resuspended in 1 ml CHCl₃/MeOH (2:1) and shaken (30 min, 4 °C). After centrifugation (5 min, 20,000g, 2 °C), the organic phase was transferred to a new tube and evaporated in a vacuum concentrator. The pellet was dissolved in SDS sample buffer.

For reconstitution, 2 mg of BJ3505 vacuoles were diluted fivefold with PSK buffer, re-isolated (10,000g, 2 °C, 3 min) and solubilized (3 min, 4 °C) at 1 mg ml⁻¹ in 1% Triton X-100 in buffer P (PSK buffer with 2 mM MgCl₂, 100 µM CaCl₂, 10% glycerol, 1 mM PMSF). The lysate was centrifuged (20,000g, 2 °C, 3 min) and the V-ATPases were immunoprecipitated using 10 µg each of monoclonal antibodies to Vph1 and Vma2 adsorbed to 100 µl bed volume Protein G Sepharose. The beads were washed twice with 0.5% Triton X-100 in buffer P, once with 0.1% Triton X-100 in buffer P, once with buffer P, sedimented, resuspended in 500 µl of vacuolar lipids (see below) and 500 µl of 20 mg ml⁻¹ asolectine (Sigma) in chloroform/methanol (1:1) and shaken (15 min, 4 °C). After centrifugation (1 min, 10,000g, room temperature), the organic phase was collected. 1.5 ml of extract was diluted with 1 ml buffer R (100 mM Tris–HCl pH 7.4, 100 mM choline chloride, 50 mM KCl, 100 µM EGTA) under vigorous vortexing. Calmodulin (100 ng ml⁻¹) could be added to buffer R. The solvent was evaporated under N₂. After adding 100 µM MgCl₂, the vesicles were centrifuged (TLA45 rotor, 10 min, 125,000g, 20 °C), resuspended in 1 ml of assay buffer (100 mM Tris–HCl pH 7.4, 150 mM KCl, 100 µM EGTA) and centrifuged again as before. The pellet was resuspended in 200 µl of assay buffer. Alternatively, flotation in a step gradient with 8% and 4% Ficoll in PS and 1–2 ml of assay buffer was used (SW41, 45 min, 150,000g, 20 °C). Vesicles from the assay buffer phase were diluted and pelleted as above.

For endogenous lipids, 10 mg frozen vacuoles was diluted with 50 ml of 10 mM Tris–HCl pH 7.4, 1 mM EDTA, and centrifuged (10 min, 15,000g, 4 °C). The pellets were resuspended in 2 ml of Tris–HCl pH 7.4, 1 mM EDTA, diluted with 20 ml of CHCl₃/MeOH (1:1), rotated (30 min, 4 °C) and centrifuged (10 min, 5,000g, 4 °C). The super-

nanat was extracted with two volumes of water and centrifuged (30 min, 15,000g, 4 °C). The interphase (proteolipids) and upper phase were discarded. The lower phase was used as the source of endogenous lipids. Proteins were not detectable in this extract.

Choline release was assayed in a microplate luminometer (Berthold). Samples contained 25 µl of assay buffer, 20 µl of vesicles, 5 µl of enzyme mix (50 µl 250 U ml⁻¹ choline oxidase in 100 mM Tris–HCl pH 7.4, 150 mM KCl; 50 µl of 10 mM luminol in 200 mM Tris–HCl pH 8.6; 25 µl of 2 mg ml⁻¹ horseradish peroxidase in 100 mM Tris–HCl pH 7.4, 150 mM KCl), and 200 µM A23187 (10 mM stock in DMSO). CaCl₂ was injected to 10 mM and luminescence was measured over time. Ten millimolar Ca²⁺ and the ionophore are probably necessary to generate a sufficiently high internal calcium concentration and membrane potential^{28,30}. The required Ca²⁺ concentration could be lowered significantly if a K⁺ diffusion potential was generated (not shown).

H⁺-pump activity of vacuoles

The V1 sector was extracted by modifying vacuole isolation: 100 mM NaSCN, 5 mM MgCl₂ and 5 mM ATP were added to the 15% Ficoll phase used for the flotation gradient. The 8% Ficoll solution contained 5 mM MgCl₂ and 5 mM ATP only. The incubation with DEAE dextran was extended (3 min, 30 °C) and the dextran concentration increased by 50%. Ultracentrifugation was reduced to 60 min.

For assay of pump activity, 100-µl fusion reactions were incubated at 25 °C with 0.5 µM BCECF-dextran (70K; Molecular Probes). The change of fluorescence was measured at an excitation of 490 nm and an emission of 520 nm in a Perkin Elmer LS50B fluorometer. Initial pump activity was determined from the slope of the linear change of signal during the first 5 min of incubation. The system was calibrated by measuring fluorescence of 0.5 µM BCECF-dextran in PSK buffer with 0.5 mM MnCl₂ buffered to pH 6.2–7.4.

For assay of H⁺ gradient, 200-µl fusion reactions were incubated with 200 µM quinacrine (10 min, 27 °C), diluted with 1 ml ice-cold PSK buffer, re-isolated (2 min, 20,000g) and solubilized in 1 ml 0.4% Triton X-100. A₄₃₀ was measured.

Immunoprecipitations

Vacuoles from a 1-ml fusion reaction were pelleted (2 min, 20,000g, 2 °C), solubilized for 3 min in buffer I (0.5% Triton X-100, 0.5 mM MnCl₂, 100 µM CaCl₂, 1 mM PMSF in PSK buffer) and centrifuged (5 min, 100,000g, 2 °C). The supernatant was supplemented with 30–60 µg of antibodies and 25 µl of Protein G Sepharose and shaken (45 min, 4 °C). The beads were washed three times with buffer I and suspended in SDS sample buffer. For calmodulin precipitations, anti-calmodulin IgGs were coupled to carbolink resin (Pierce) according to the manufacturer's instructions.

Received 10 August; accepted 29 November 2000.

- Jahn, R. & Südhof, T. C. Membrane fusion and exocytosis. *Annu. Rev. Biochem.* **68**, 863–911. (1999).
- Pfeffer, S. R. Transport vesicle targeting: tethers before SNAREs. *Nature Cell Biol.* **1**, 17–19 (1999).
- Mayer, A. Membrane fusion: SNAREs only? *Curr. Opin. Cell Biol.* **11**, 447–452 (1999).
- Lindau, M. & Almers, W. Structure and function of fusion pores in exocytosis and ectoplasmic membrane fusion. *Curr. Opin. Cell Biol.* **7**, 509–517 (1995).
- Chernomordik, L. V. & Zimmerberg, J. Bending membranes to the task: structural intermediates in bilayer fusion. *Curr. Opin. Struct. Biol.* **5**, 541–547 (1995).
- Hanson, P. I. *et al.* Structure & conformational changes in NSF and its membrane receptor complexes visualized by quick-freeze/deep-etch electron microscopy. *Cell* **90**, 523–535 (1997).
- Skehel, J. J. & Wiley, D. C. Coiled coils in both intracellular vesicle and viral membrane fusion. *Cell* **95**, 871–874 (1998).
- Weber, T. *et al.* SNAREpins: Minimal machinery for membrane fusion. *Cell* **92**, 759–772 (1998).
- Otter-Nilsson, M. *et al.* Cytosolic ATPases, p97 and NSF, are sufficient to mediate rapid membrane fusion. *EMBO J.* **18**, 2074–2083 (1999).
- Brügger, B. *et al.* Putative fusogenic activity of NSF is restricted to a lipid mixture whose coalescence is also triggered by other factors. *EMBO J.* **19**, 1272–1278 (2000).
- Coorssen, J. R., Blank, P. S., Tahara, M. & Zimmerberg, J. Biochemical and functional studies of cortical vesicle fusion: The SNARE complex and Ca²⁺ sensitivity. *J. Cell Biol.* **143**, 1845–1857 (1998).
- Chen, Y. A. *et al.* SNARE complex formation is triggered by Ca²⁺ and drives membrane fusion. *Cell* **97**, 165–174 (1999).
- Xu, T. *et al.* Inhibition of SNARE complex assembly differentially affects kinetic components of exocytosis. *Cell* **99**, 713–722 (1999).
- Ungermann, C., Sato, K. & Wickner, W. Defining the functions of *trans*-SNARE pairs. *Nature* **396**, 543–548 (1998).
- Ungermann, C., Wickner, W. & Xu, Z. Y. Vacuole acidification is required for *trans*-SNARE pairing, LMA1 release, and homotypic fusion. *Proc. Natl Acad. Sci. USA* **96**, 11194–11199 (1999).
- Peters, C. & Mayer, A. Ca²⁺/calmodulin signals the completion of docking and triggers a late step of vacuole fusion. *Nature* **396**, 575–580 (1998).
- Peters, C. *et al.* Control of the terminal step of membrane fusion by protein phosphatase 1. *Science* **285**, 1084–1087 (1999).
- Pryor, P. R. *et al.* The role of intraorganellar Ca²⁺ in late endosome–lysosome heterotypic fusion and in the reformation of lysosomes from hybrid organelles. *J. Cell Biol.* **149**, 1053–1062 (2000).
- Colombo, M. I., Beron, W. & Stahl, P. D. Calmodulin regulates endosome fusion. *J. Biol. Chem.* **272**, 7707–7712 (1997).
- Holroyd, C., Kistner, U., Annaert, W. & Jahn, R. Fusion of endosomes involved in synaptic vesicle recycling. *Mol. Biol. Cell* **10**, 3035–3044 (1999).
- Porat, A. & Elazar, Z. Regulation of intra-Golgi membrane transport by calcium. *J. Biol. Chem.* **275**, 29233–29237 (2000).
- Mayer, A. & Wickner, W. Docking of yeast vacuoles is catalysed by the ras-like GTPase Ypt7p after symmetric priming by Sec18p (NSF). *J. Cell Biol.* **136**, 307–317 (1997).
- Ungermann, C., Nichols, B. J., Pelham, H. R. & Wickner, W. A vacuolar v-t-SNARE complex, the predominant form in vivo and on isolated vacuoles, is disassembled and activated for docking and fusion. *J. Cell Biol.* **140**, 61–69 (1998).
- Harteringer, J., Stenius, K., Högemann, D. & Jahn, R. 16-BAC/SDS-PAGE: a two-dimensional gel electrophoresis system suitable for the separation of integral membrane proteins. *Anal. Biochem.* **240**, 126–133 (1996).
- Cohen, A., Perzov, N., Nelson, H. & Nelson, N. A novel family of yeast chaperons involved in the distribution of V-ATPase and other membrane proteins. *J. Biol. Chem.* **274**, 26885–26893 (1999).
- Stevens, T. H. & Forgac, M. Structure, function and regulation of the vacuolar ATPase. *Annu. Rev. Cell Dev. Biol.* **13**, 779–808 (1997).
- Kane, P. M. & Parra, K. Assembly and regulation of the yeast vacuolar H⁺-ATPase. *J. Exp. Biol.* **203**, 81–87 (2000).
- Dunant, Y. & Israël, M. In vitro reconstitution of neurotransmitter release. *Neurochem. Res.* **23**, 709–718 (1998).
- Israël, M., Morel, N. & Lesbats, B. Evidence for an association of the 15-kDa proteolipid of mediophore with a 14-kDa polypeptide. *J. Neurochem.* **57**, 2047–2053 (1991).
- Israël, M., Meunier, F. M., Morel, N. & Lesbats, B. Calcium-induced desensitization of acetylcholine release from synaptosomes or proteoliposomes equipped with mediophore, a presynaptic membrane protein. *J. Neurochem.* **49**, 975–982 (1987).
- Mayer, A., Wickner, W. & Haas, A. Sec18p (NSF) driven release of Sec17p (alpha-SNAP) can precede docking and fusion of yeast vacuoles. *Cell* **85**, 83–94 (1996).
- Finbow, M. E. & Harrison, M. A. The vacuolar ATPase: a universal proton pump of eukaryotes. *Biochem. J.* **324**, 697–712 (1997).
- Weber, T. *et al.* SNAREpins are functionally resistant to disruption by NSF and alpha-SNAP. *J. Cell Biol.* **149**, 1063–1072 (2000).
- Umemoto, N., Yoshihisa, T., Hirata, R. & Anraku, Y. Roles of the Vma3 gene product, subunit c of the vacuolar membrane H⁺-ATPase on vacuolar acidification and protein transport. *J. Biol. Chem.* **265**, 18447–18453 (1990).
- Klionsky, D. J., Nelson, H. & Nelson, N. Compartment acidification is required for efficient sorting of proteins to the vacuole in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **267**, 3416–3422 (1992).
- Morano, K. A. & Klionsky, D. J. Differential effects of compartment deacidification on the targeting of membrane and soluble proteins to the vacuole in yeast. *J. Cell Sci.* **107**, 2813–2824 (1994).
- Yamashiro, C. T. *et al.* Role of vacuolar acidification in protein sorting and zymogen activation: a genetic analysis of the yeast vacuolar proton translocating ATPase. *Mol. Cell. Biol.* **10**, 3737–3749 (1990).
- Kibble, V. A. & Burgoyne, R. D. Calmodulin increases the initial rate of exocytosis in adrenal chromaffin cells. *Eur. J. Physiol.* **431**, 464–466 (1996).
- Chen, Y., Duvvuri, V., Schulman, H. & Scheller, R. H. Calmodulin and protein kinase C increase Ca²⁺-stimulated secretion by modulating membrane-attached exocytic machinery. *J. Biol. Chem.* **274**, 26469–26476 (1999).
- Bennett, M. K., Calakos, N., Kreiner, T. & Scheller, R. H. Synaptic vesicle membrane proteins interact to form a multimeric complex. *J. Cell Biol.* **116**, 761–775 (1992).
- Galli, T., McPherson, P. S. & De Camilli, P. The V₀ sector of the V-ATPase, synaptobrevin, and synaptophysin are associated on synaptic vesicles in a Triton X-100-resistant, freeze-thawing sensitive complex. *J. Biol. Chem.* **271**, 2193–2198 (1996).
- Harvey, W. R. & Wiczkorek, H. Animal plasma membrane energization by chemiosmotic H⁺ V-ATPases. *J. Exp. Biol.* **200**, 203–216 (1997).
- Conchon, S., Cao, X., Barlowe, C. & Pelham, H. R. B. Got1p and Sft2p: membrane proteins involved in traffic to the Golgi complex. *EMBO J.* **18**, 3934–3946 (1999).
- Zimmerberg, J., Vogel, S. & Chernomordik, L. V. Mechanisms of membrane fusion. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 433–466 (1993).
- Albillos, A. *et al.* The exocytotic event in chromaffin cells revealed by patch amperometry. *Nature* **389**, 509–512 (1997).
- Henkel, A. & Betz, W. J. Staurosporine blocks evoked release of FM1-43 but not acetylcholine from frog motor nerve terminals. *J. Neurosci.* **15**, 8246–8258 (1995).
- Scepek, S., Coorssen, J. R. & Lindau, M. Fusion pore expansion in horse eosinophils is modulated by Ca²⁺ and protein kinase C via distinct mechanisms. *EMBO J.* **17**, 4340–4345 (1998).
- Plattner, H. & Knoll, G. in *Signal Transduction During Biomembrane Fusion* (ed. D. H. O'Day) 19–46 (Academic, San Diego, 1993).
- García-Segura, L. M., Müller, D. & Dunant, Y. Increase in the number of presynaptic large intramembrane particles during synaptic transmission at the Torpedo nerve–electroplaque junction. *Neuroscience* **19**, 63–78 (1986).
- Chernomordik, L., Kozlov, M. M. & Zimmerberg, J. Lipids in biological membrane fusion. *J. Membr. Biol.* **146**, 1–14 (1995).
- Shevchenko, A. *et al.* Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl Acad. Sci. USA* **93**, 14440–14445 (1996).

Acknowledgements

We thank D. Gallwitz, P. Kane, R. Piper and M. Harrison for plasmids and strains; C. Baradov for assistance; and the Boehringer Ingelheim Foundation and Deutsche Forschungsgemeinschaft (SFB446) for support.

Correspondence and requests for materials should be addressed to A.M. (e-mail: andreas.mayer@tuebingen.mpg.de).

Rapid collisional evolution of comets during the formation of the Oort cloud

S. Alan Stern* & Paul R. Weissman†

* Space Studies Department, Southwest Research Institute, Boulder, Colorado 80302, USA

† Earth and Space Sciences Division, Jet Propulsion Laboratory, Pasadena, California 91109, USA

The Oort cloud¹ of comets was formed by the ejection of icy planetesimals from the region of giant planets—Jupiter, Saturn, Uranus and Neptune—during their formation². Dynamical simulations^{3,4} have previously shown that comets reach the Oort cloud only after being perturbed into eccentric orbits that result in close encounters with the giant planets, which then eject them to distant orbits about 10^4 to 10^5 AU from the Sun (1 AU is the average Earth–Sun distance). All of the Oort cloud models constructed until now simulate its formation using only gravitational effects; these include the influence of the Sun, the planets and external perturbers such as passing stars and Galactic tides. Here we show that physical collisions between comets and small debris play a fundamental and hitherto unexplored role throughout most of the ejection process. For standard models of the protosolar nebula (starting with a minimum-mass nebula) we find that collisional evolution of comets is so severe that their erosional lifetimes are much shorter than the timescale for dynamical ejection. It therefore appears that collisions will prevent most comets escaping from most locations in the region of the giant planets until the disk mass there declines sufficiently that the dynamical ejection timescale is shorter than the collisional lifetime. One consequence is that the total mass of comets in the Oort cloud may be less than currently believed.

We consider a solar nebula consisting of the four giant planets and a mass of residual solid material, smoothly distributed between 5 and 50 AU, equal to $\eta \times 50M_{\oplus}$: here η is a scalar and $M_{\oplus}$ is the mass of the Earth; $50M_{\oplus}$ is the mass derived by extrapolating an estimate of the mass of the solid material in the giant planets⁵ to a nebula extending outwards to 50 AU. As with published models of the formation of the Oort cloud, we ignore the presence of gas. We adopt the standard (though idealized) assumption that the nebula is distributed in an axially symmetric disk with surface mass density falling off with heliocentric distance, a , as $a^{-3/2}$. If giant-planet formation is efficient (that is, if most of the residual material is accreted into the giant planets), then $\eta < 1$. However, η has been estimated⁶ to be probably ~ 3 – 10 .

To begin our assessment of the role of physical collisions, we first show that the disk is so collisional that momentum transfer during collisions affects the orbits of the objects in the disk. A good criterion for this⁷ is that an object's dynamics will be severely affected by the transfer of momentum when it has encountered one-third to one-half of its own mass from the impacts. Therefore, in Fig. 1 we show the timescale, $T_{\text{crit}}(r_c)$, required for a typical comet (radius $r_c = 1$ km and density $\rho_c = 1 \text{ g cm}^{-3}$), orbiting wholly within the nebular disk at approximately constant heliocentric distance a , to collide against one-half its own mass, for three assumed values of $\eta = 10$, 1 and 0.1. This timescale is the ratio of half the mass of an orbiting comet, m_c , to the rate $\dot{m}_c$ at which it encounters mass from collisions with other objects in the ensemble. This timescale can be derived under the assumptions of negligible gravitational focusing by the comets, and mean orbital inclinations $\langle i \rangle$ at any given heliocentric distance such that the comets are wholly embedded in the disk. If we assume that the mean orbital inclination $\langle i \rangle = (1/2)\langle e \rangle$, $\langle e \rangle \ll 1$ and $2a\langle e \rangle \gg 2r_c$, then $T_{\text{crit}} = m_c / (2\dot{m}_c) = [2ar_c\rho_c / (3\sqrt{3}\nu_k^2\Sigma_0)](a/a_0)^2$, where a_0 is a normalization distance (for

example, 1 AU), a is the heliocentric distance of interest, ν_k^0 is the keplerian circular orbit velocity at a_0 , $\langle e \rangle$ is the mean random eccentricity and Σ_0 is the surface mass density of solid bodies at a_0 with radii $r < r_c$.

We can compare T_{crit} to the overplotted red curves in Fig. 1, which represent the time required for most test particles in recent dynamical simulations to reach planet-crossing orbits, as a function of their starting semi-major axes^{8,9}. Once objects become planet-crossing, they are rapidly ejected from the system. Notice that under the standard model assumptions outlined above, comets will be struck by a mass comparable to their own on timescales of $10^{4.5-6.5}$ years, depending on their heliocentric location, until η declines below about 1 to 0.5. Figure 1 further shows that this collisional timescale is significantly shorter than the time it takes most test particles on initially near-circular orbits to encounter a giant planet.

The effect of such impacts will be to damp orbital eccentricities and inclinations. This will, in turn, prevent objects from reaching the giant-planet-crossing orbits required to achieve the strong scattering events that ultimately eject them from the region of the giant planets, while prolonging the time available for collisions. This calls into question results obtained previously from purely gravitational simulations of the formation of the Oort cloud, because such models allow orbital eccentricities, inclinations and semi-major axes to evolve unimpeded during the ejection process.

We next consider the role of energetic collisions that cause objects to lose mass (that is, to erode) while still in the region of the giant planets. The mean random impact speed in collisions in a three-dimensional disk between small bodies that do not themselves have significant escape speeds is closely approximated by $\sqrt{3}\langle e \rangle \nu_k$, where $\langle e \rangle$ is their mean random orbital eccentricity and ν_k is the local keplerian circular orbit speed. We used a theory developed¹⁰ to calculate the critical mean eccentricity, $\langle e^* \rangle$, above which an object of specified strength and size (that is, gravitational binding energy) will, if struck, suffer net erosion rather than accretion. We found that across the region of the giant planets, even for the assumption

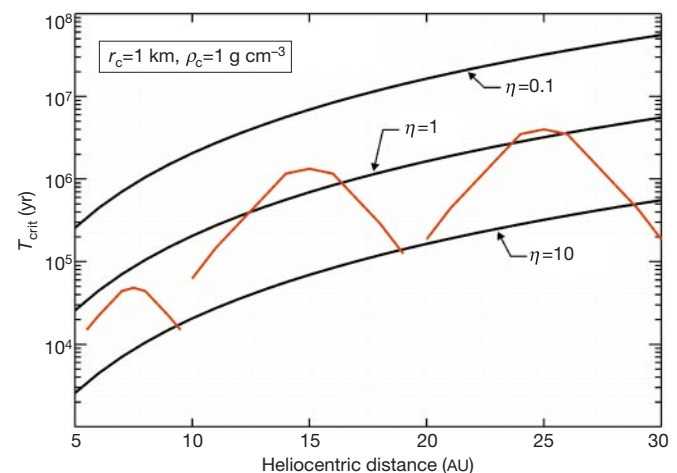


Figure 1 Estimated time required for cometary nuclei to collide with 50% of their own mass in solids in the solar nebula as a function of distance from the Sun. Black curves, the time required for a typical comet, with radius $r_c = 1$ km and density $\rho_c = 1 \text{ g cm}^{-3}$, orbiting wholly within the nebular disk on a low-eccentricity, low-inclination orbit at a , to collide with a cumulative mass equal to 50% of its own mass for three values of η : 10, 1 and 0.1. For this calculation we assumed $\Sigma_0 = 2 \text{ g cm}^{-2}$ at $a_0 = 5 \text{ AU}$ (a canonical minimum-mass solar nebula⁵), scaling the disk surface mass density as $a^{-3/2}$. The timescales for other assumed cometary radii and densities scale linearly. Red curves, the timescale for 75% of all these test particles to reach planet-crossing orbits^{8,9}, once objects become planet-crossing, they are rapidly ejected from the system. For simplicity, we have plotted smoothed versions of these dynamical timescales. The dynamical results^{8,9} were generated from simulations beginning with one-third of the particles at each distance with $\langle e \rangle > 0.10$, and nine-tenths of the particles at each distance with $\langle e \rangle > 0.01$.

of mechanically strong comets with relatively large radii of 10 km, $\langle e^* \rangle > 5 \times 10^{-3}$, results in erosion when collisions occur; $\langle e^* \rangle > 5 \times 10^{-4}$ results in erosive mass loss for weak comets with more typical radii of 1 km. These results indicate that the clearing of the zone of the giant planets, which dynamical models indicate required eccentricities up to 0.3 or more, initiated a period of highly erosive mass loss among the small bodies being cleared.

We then used a time-dependent collisional simulation¹¹ to estimate erosional lifetimes for typical 1-km-radius objects in a disk-like solar nebula. This statistical, particle-in-a-box code is based on Wetherill's "moving bin scheme"¹², and computes the rates and effects of collisions across a multi-zone particle disk using modern collisional scaling laws¹³ to evolve the size–frequency distribution as a function of time. We performed simulations assuming a three-dimensional disk from 5 to 50 AU, consisting of $50M_{\oplus}$ (that is, $\eta = 1$) of comets and small debris, with a surface mass density that falls off as $a^{-3/2}$. The starting population is assumed to be described by a standard collisional equilibrium power law size distribution¹⁴. We assume the orbits of the comets are contained wholly within the planetesimal disk before they become planet crossing. We again assumed $\langle i \rangle = (1/2)\langle e \rangle$, which closely obtains for an ensemble in dynamical equilibrium. We also assumed that this ensemble is gas free and consists of bodies with radii ranging from 1 cm to 200 km.

The curves in Fig. 2 show that erosion rates are greatest in the Jupiter–Saturn zone, but are still substantial in the Uranus–Neptune zone, even for an $\langle e \rangle$ of only 0.01. Comparing these rates with the dynamical lifetimes of test particles in these regions (which are of the order of $10^{4.5}$ years in the Jupiter–Saturn zone, 10^6 years in the Saturn–Uranus zone, and $10^{6.5}$ years in the Uranus–Neptune zone^{8,9,15}), Fig. 2 indicates that the erosional timescales for kilometre-sized objects in the giant-planets region are generally shorter than their dynamical lifetimes as long as $\eta \approx 0.3$. Further, the escape of comets to the Oort cloud from the region of any given giant planet will be efficient only after the disk evolves to a state in which the local collisional mean free path is long enough to permit objects the size of typical comets to escape largely intact. This can be accomplished either by the collisional grinding down of many objects to dust particle sizes where radiation-pressure-driven transport becomes important, or by the growth of a comparatively small number of larger bodies which trap a large fraction of the disk's mass.

We also ran the model for comets of radius 10 km and 30 km using the same four test cases, and found that objects up to ~ 10 km radius were rapidly eroded away by collisions with smaller objects before they could be ejected to the comparative safety of the Oort cloud (where collision rates are negligible¹⁶). When $\eta = 1$, only objects significantly larger than 10 km could avoid erosion or destruction before escape to the Oort cloud.

Our results show that collisions play a fundamental role in shaping the dynamical and physical evolution of comets and smaller debris between the giant planets, up to the time that the mass of this material declines to a small fraction of the solid material in the minimum-mass solar nebula. Given that simulations^{4,18} of Oort cloud formation find that only a small fraction—about 3% to 10% of the ejected material from the giant-planets region now resides in the Oort cloud (most of the remainder having been lost from the Solar System altogether), our results also suggest that the giant-planets region could only contribute ~ 0.6 – $2 M_{\oplus}$ of material to the Oort cloud. This is significantly less than conventional estimates of the mass of the Oort cloud, which typically range from $10M_{\oplus}$ to $40M_{\oplus}$ (ref. 18).

Such a low-mass Oort cloud could be reconciled with previous mass estimates if either: (1) the de-biasing of the observed long-period comet flux for observational incompleteness overestimates the cometary flux in the planetary region^{18,19}, or (2) cometary radii (which are notoriously difficult to determine) or cometary densities have been overestimated. While a low-mass Oort cloud would not

be required if the emplacement efficiency of comets into the cloud were significantly higher than the 3–10% estimated in recent simulations^{4,17,20}, it could resolve the angular-momentum dilemma implicit in the giant planets ejecting many tens to several hundred Earth masses of material^{14,21} to build a conventional Oort cloud of a few tens of Earth masses.

Whether or not the Oort cloud is indeed far lower in mass than conventional models suggest, the pervasiveness of collisional evolution demonstrates that the clearing of the giant-planets region and Oort cloud formation cannot have occurred in the manner represented by the purely gravitational simulations performed to date. Because our work involved several idealized assumptions, it represents only a first step toward the more complete modelling required. Additional factors that need to be investigated in future Oort cloud formation studies include: (1) the possibility that the planets (or large planetary embryos) in the giant-planets region scattered comets to high inclinations (that is, greater than the nebular disk thickness), thereby lengthening collisional lifetimes, and (2) that massive bodies may have migrated through the disk²², shortening the dynamical clearing timescale and reducing the effects of collisions. There may also exist dynamical 'escape valves', that is, gaps that rapidly opened up in the nebula, allowing comets to be ejected from some regions before they could be collisionally eroded.

One 'escape valve' is already suggested by the curves in Fig. 2. This is the region between 28 and 33 AU, from which objects are apparently ejected on timescales shorter than their erosional lifetimes. Realistic models of the solar nebula that account for the accretion inefficiency inherent in building Neptune suggest that this region alone may have originally contained $30M_{\oplus}$ or more of solids²³. Clearly, no more than half of this material was incorporated

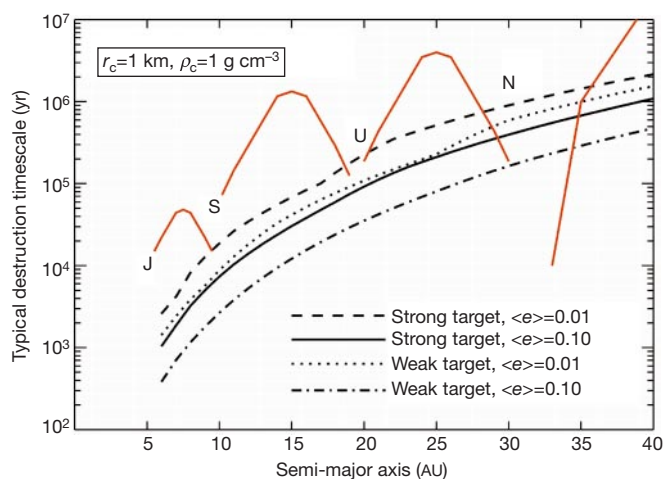


Figure 2 Estimated erosion lifetimes for icy planetesimals, embedded in a disk-like, primordial solar nebula between 5 and 50 AU from the Sun. Black curves, the timescale for a typical comet with radius $r_c = 1$ km and density $\rho_c = 1$ g cm⁻³, orbiting wholly within the nebular disk on a low-inclination, low eccentricity orbit at near-constant distance a to be destroyed by collisions. Red curves, 75% time required for all these test particles to reach planet-crossing orbits^{8,9}; once objects become planet-crossing, they are rapidly ejected from the system. Results are shown for two values of the initial eccentricity, 0.01 and 0.10, and for two target (planetesimal) strengths: (1) a 'weak' tensile strength of 3×10^3 erg cm⁻³ (typical of that estimated from the observed tidal disruption of comets passing near the Sun or Jupiter²⁷), and (2) a 'strong' tensile strength of 3×10^6 erg cm⁻³ (more typical of refractory solids¹³). The combinations of assumed tensile strengths and eccentricities that we chose span a wide region of the plausible parameter space. The dynamical results^{8,9}, shown in red, were adopted from the same simulations used for Fig. 1. The heliocentric distances of the four giant planets—Jupiter (J), Saturn (S), Uranus (U) and Neptune (N)—are shown here for reference. Higher values of $\langle e \rangle$ increase the erosion rate by increasing impact energies. These results show that erosional lifetimes are shorter than the dynamical ejection times in all four giant-planet zones.

into Neptune. Following this scenario further by adopting the 7–10% Oort cloud dynamical emplacement efficiency found previously at this heliocentric distance⁴, we estimate that up to $\sim 1.5M_{\odot}$ of additional material may have been ejected from this region to the Oort cloud early on. Other distant ‘escape valves’, perhaps including mean-motion resonances in the Kuiper belt, may have also contributed to this mass flux. Including such sources could result in a total Oort cloud mass up to $\sim 3.5M_{\odot}$, with a significant or dominant fraction of the comets in this cloud originating in the Neptune and Kuiper belt regions. Although this is still low compared to recent estimates of the mass of the Oort cloud, we speculate that such a distant source region for Oort cloud comets could reconcile recent findings^{24–26} that all Oort cloud comets with measured thermal history constraints formed at significantly lower temperatures than conventional models of the cloud have predicted. Both this possibility, as well as our main results given above, suggest the need for a general reappraisal of Oort cloud formation models, using coupled collisional–dynamical simulations. □

Received 28 September; accepted 13 December 2000.

- Oort, J. H. The structure of a cloud of comets surrounding the solar system and a hypothesis concerning its structure. *Bull. Astron. Inst. Neth.* **11**, 91–110 (1950).
- Kuiper, G. P. in *Astrophysics: A Topical Symposium* (ed. Hynek, J. A.) 357–424 (McGraw-Hill, New York, 1951).
- Gladman, B. & Duncan, M. J. Fates of minor bodies in the outer solar system. *Astron. J.* **100**, 1680–1693 (1990).
- Dones, L., Levison, H. F., Duncan, M. J. & Weissman, P. R. Formation of the Oort cloud revisited. *Bull. Am. Astron. Soc.* **32**, 1060 (2000).
- Wuchterl, G., Guillot, T. & Lissauer, J. J. in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) 1081–1110 (Univ. Arizona Press, Tucson, 2000).
- Lissauer, J. J. Time scales for planetary accretion and the structure of the protoplanetary disk. *Icarus* **69**, 249–267 (1987).
- Pollack, J. B., Burns, J. A. & Tauber, M. E. Gas drag in primordial circumplanetary envelopes: A mechanism for satellite capture. *Icarus* **37**, 587–611 (1979).
- Grazier, K., Newman, W. L., Kaula, W. M. & Hyman, J. M. Dynamical evolution of planetesimals in the outer solar system I. *Icarus* **140**, 341–252 (1999).
- Grazier, K., Newman, W. L., Kaula, W. M. & Hyman, J. M. Dynamical evolution of planetesimals in the outer solar system II. *Icarus* **140**, 353–368 (1999).
- Stern, S. A. Collisional timescales and the architecture of the ancient, massive Kuiper disk. *Astron. J.* **112**, 1203–1210 (1996).
- Stern, S. A. & Colwell, J. E. Collisional erosion in the Edgeworth–Kuiper belt. *Astrophys. J.* **490**, 879–884 (1997).
- Wetherill, G. W. Comparison of analytical and physical modelling of planetesimal accumulation. *Icarus* **88**, 336–354 (1990).
- Housen, K. R. & Holsapple, K. A. On the fragmentation of asteroids and planetary satellites. *Icarus* **84**, 226–253 (1990).
- Dohnanyi, J. S. Collisional model of asteroids and their debris. *J. Geophys. Res.* **74**, 2531–2554 (1969).
- Holman, M. J. & Wisdom, J. Dynamical stability in the outer solar system and the delivery of short period comets. *Astron. J.* **105**, 1987–1999 (1993).
- Stern, S. A. Collisions in the Oort cloud. *Icarus* **73**, 499–505 (1988).
- Duncan, M. J., Quinn, T. & Tremaine, S. The formation and extent of the solar system Oort cloud. *Astron. J.* **94**, 1330–1338 (1987).
- Weissman, P. R. in *Completing the Inventory of the Solar System* (eds Rettig, T. W. & Hahn, J. M.) 265–288 (ASP Conf. Ser. 107, Astronomical Society of the Pacific, San Francisco, 1996).
- Everhart, E. Comet discoveries and observational selection. *Astron. J.* **72**, 716–726 (1967).
- Bailey, M. E. & Stagg, C. R. Cratering constraints on the inner Oort cloud and implications for cometary origins. *Mon. Not. R. Astron. Soc.* **235**, 1–32 (1988).
- Safronov, V. S. Kuiper prize lecture: Some problems in the formation of the planets. *Icarus* **94**, 260–271 (1991).
- Thommes, E. W., Duncan, M. J. & Levison, H. F. The formation of Uranus and Neptune in the Jupiter–Saturn region of the Solar System. *Nature* **402**, 635–638 (1999).
- Lissauer, J. J., Pollack, J. B., Wetherill, G. W. & Stevenson, D. J. in *Neptune and Triton* (eds Cruikshank, D. P. & Matthews, M. S.) 37–108 (Univ. Arizona Press, Tucson, 1995).
- Mumma, M. J. in *From Stardust to Planetesimals* (eds Pendleton, Y. J. & Tielens, A. G. G. M.) 369–396 (ASP Conf. Ser. 122, Astronomical Society of the Pacific, San Francisco, 1997).
- Blake, G. A., Hogerheijde, M. R., Gurwell, M. A. & Muhleman, D. O. Sublimation from icy jets as a probe of the interstellar volatile content of comets. *Nature* **398**, 213–216 (1999).
- Stern, S. A. *et al.* The discovery of argon in comet C/1995 O1 (Hale-Bopp). *Astrophys. J.* **545**, L169–172 (2000).
- Sekanina, Z. in *Comets* (ed. Wilkening, L. L.) 251–287 (Univ. Arizona Press, Tucson, 1982).

Acknowledgements

We thank M. Duncan, K. Grazier and H. Levison for discussions, and D. Durda, W. Merline, A. Morbidelli, W. Ward and most particularly L. Dones for comments on this manuscript. This work was performed in part at the Jet Propulsion Laboratory under contract with NASA, and was supported by the NASA Planetary Geology and Geophysics and NASA Origins Programs.

Correspondence and requests for materials should be addressed to S.A.S. (e-mail: astern@swri.edu).

Two coexisting vortex phases in the peak effect regime in a superconductor

M. Marchevsky*, M. J. Higgins* & S. Bhattacharya*†

* NEC Research Institute, Princeton, New Jersey 08540, USA

† Tata Institute of Fundamental Research, Mumbai-400005, India

The critical current in the vortex phase of a type-II superconductor such as NbSe₂ displays a striking anomaly in the vicinity of the superconductor-to-normal-metal transition. Instead of going to zero smoothly, it rebounds to a sharp and pronounced maximum, just before vanishing at the transition. This counter-intuitive phenomenon, known as the peak effect^{1–3}, has remained an unsolved problem for 40 years. Here we use a scanning a.c. Hall microscope to visualize the real-space distribution of the critical current in NbSe₂. We show that in the peak-effect regime two distinct vortex-matter phases with intrinsically different pinning strengths coexist on a macroscopic scale. The composition of the two-phase mixture and the transformation of one phase into another are responsible for the history effects^{4–6} and anomalous voltage response^{4,5} commonly seen when external parameters such as temperature, magnetic field or transport current are varied. We argue that the observed phase coexistence is, in fact, the hallmark of a disorder-driven non-thermal phase transition.

Our experiments were performed using a variant of scanning Hall probe microscopy⁷. The Hall probe assembly, part of a purpose-built low-temperature scanning probe microscope, is shown schematically in Fig. 1a inset. The microscope provides a map of the local induction, with a spatial resolution defined by the active area of the Hall sensor in the *x*–*y* plane. In our system we apply a small a.c. magnetic excitation field and detect the local a.c. magnetic response with the Hall sensor. In the case of vortices in superconductors, the a.c. field transmitted through the sample and detected at its surface (*B*_{ac}) is directly related⁸ to the mobility of the vortices, averaged over the sample thickness. For example, an increase in the pinning of the vortices, and in turn an increase in the critical current density *j*_c, will reduce the modulation amplitude *B*_{ac} detected by the Hall sensor. The peak in *j*_c is thus seen as a minimum in *B*_{ac}. These studies were conducted on single-crystal samples of the conventional superconductor 2H-NbSe₂ where the peak effect is readily detected and the occurrence of a variety of dynamic anomalies and history effects has been widely reported^{3–6}.

The Hall sensor was scanned at a distance of about 0.5 μm above the surface of the sample. The sample was cooled in an applied d.c. magnetic field of ~270 G, and images were taken with a temperature step of 5 mK. Figure 1b shows a representative set of images; the temperature corresponding to each frame is shown in Fig. 1a. The area of the Hall sensor contains about ten vortices at this field induction, and therefore individual vortices are not resolved. The plot of the temperature dependence of *B*_{ac}, averaged over the entire scanned area (Fig. 1a), clearly shows the peak effect.

As the sample is cooled from the normal state (shown as frame 1 in Fig. 1b), a sharp transition occurs to a highly pinned (blue) phase (frame 3). As the sample is further cooled, the weak-pinning (yellow) phase begins to enter the field of view from the interior of the sample (the upper right corner), and produces a stable interface with the strong-pinning phase. Upon further cooling, this phase coexistence becomes more apparent as the interface progressively invades the field of view from the upper right and moves towards the lower left corner, which is also the closest location to the sample edge. Thus the typical peak regime, between frames 3 and 8, marks a region of stable (within experimental

timescales) coexistence of a strong-pinning phase with a weak-pinning phase and a gradual transformation of one into the other. The coexistence region is about 50 mK wide. This width is much larger than the sharp transition to a highly pinned phase that occurs between frames 1 and 3, and which provides the upper limit of thermal inhomogeneity of the sample. The gradual transformation in the coexistence region is, moreover, not uniform; this is clearly seen in frame 7 where the blue region is trapped in the lower middle part, while the yellow phase goes around it.

In order to compare the local response in different spatial locations on the sample, we define regions I, II and III in the frame-set in Fig. 1b (see frame 1) and plot in Fig. 2 the evolution of average B_{ac} in these regions. The frame-averaged temperature dependence of B_{ac} from Fig. 1a is also plotted in Fig. 2 for comparison. The results show that the course of the phase transformation is position-dependent, and is influenced by the spatial inhomogeneity in the sample. This dependence leads to a complex connectivity of the two-phase coexistence region and provides an explanation of the multiple peak structures seen in, for example, Fig. 2 and reported widely in the literature⁶.

Next, we address the history-dependent effects in the peak regime by studying the behaviour in the same spatial location in the same sample for a different thermomagnetic history. In this case the sample was cooled to 5 K, and the external d.c. field was then applied (the zero-field cooled state). The temperature was increased in steps, and the images were acquired using the same procedure as in Fig. 1. In Fig. 3, the frame-averaged B_{ac} is plotted; the corresponding data from Fig. 1a are shown as a dotted line, for comparison. The imaging data show that the same sequence of events occurs, but now

in reverse order as the temperature is increased. A representative frame (marked A) taken at $T = 5.97$ K, together with the corresponding frame from the field cooling experiment (marked B), are shown in Fig. 3 inset. The strong-pinning phase appears from the left side of the frame and gradually fills up the field of view. The details, however, are quite different from the field-cooled case. The interface between the phases is not as distinct as in the field-cooled case shown above. Also, it propagates along a different path. As a result, the average B_{ac} is also different, thus manifesting history dependence in both microscopic and macroscopic scale. It is, therefore, the topology of the two-phase mixture that is primarily responsible for the history dependence in the peak-effect region.

We investigated this phase transformation for several applied d.c. magnetic fields. These measurements were done on a different sample and without scanning the sensor. Owing to a non-local relation between the current and the induction field B (described by the Biot–Savart law), the field of the currents flowing in the remote areas of the sample is always a part of the locally measured B_{ac} . This non-local contribution, indicating the global redistribution of the shielding currents, allows detection of the phase-boundary motion, even though this motion occurs at a large distance from the sensor. (In our imaging experiments, a non-local contribution, indicating the redistribution of a.c. currents outside the scanned area, is seen as a spatially uniform variation of the image amplitude.)

The results are collected in Fig. 4, where the temperature range between the initial lower-temperature ‘dip’ in B_{ac} and the last point before recovery at T_p is marked as the ‘phase coexistence’ region for

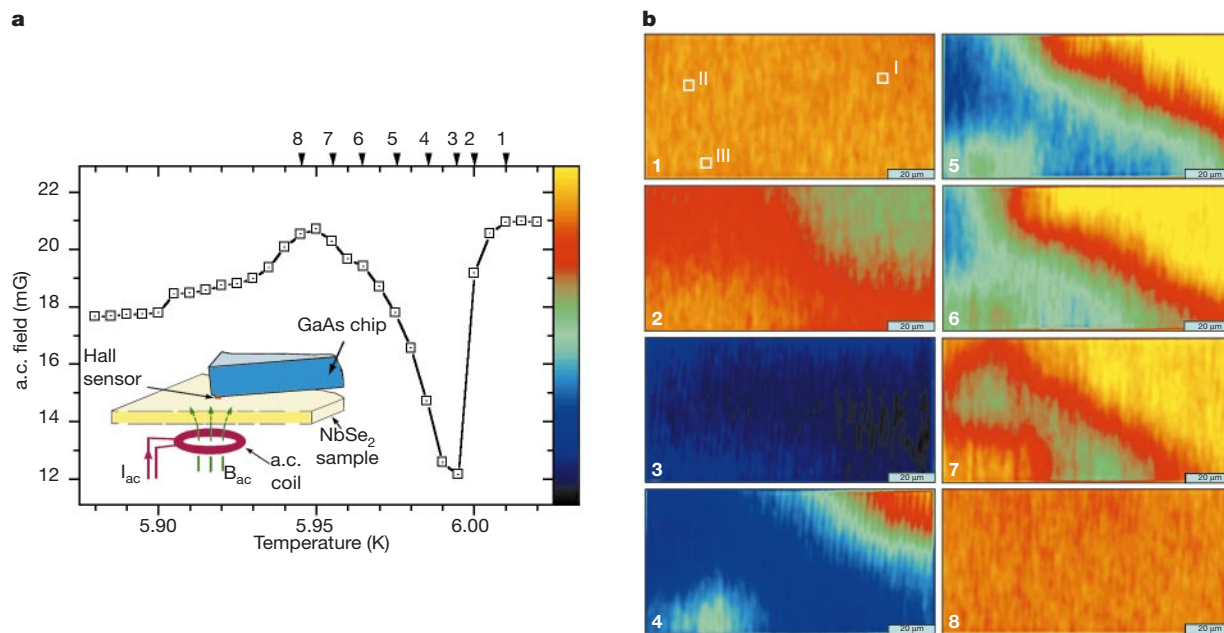


Figure 1 Imaging the peak effect. **a**, Inset: sketch of the experimental arrangement. A miniature coil (diameter, ≈ 1 mm) is placed on one side of the 2H-NbSe₂ single crystal ($8 \times 1.5 \times 0.05$ mm); an a.c. current (~ 1 mA), frequency $f = 1,377$ Hz, is applied to the coil. The a.c. magnetic field B_{ac} is measured at the opposite side of the sample with a Hall sensor, microfabricated near the corner of the 2DEG (two-dimensional electron gas) GaAs chip. The chip is mounted on the scanning probe microscope's scanning stage (not shown), and scanned over the surface of the sample in the range of $\sim 150 \mu\text{m}$ (at 6 K). Main figure: plot of B_{ac} averaged over the entire frame area, showing the peak effect. The measuring temperatures of the representative frames are indicated with arrowheads at the upper x axis of the plot. The colour bar near the y axis on the right represents the colour-coding scheme for B_{ac} used in the scanned images. **b**, A set of eight representative

frames, $130 \times 65 \mu\text{m}$, scanned at a rate of 30 min per frame in the course of a cooldown through the peak effect in an applied d.c. field of 270 G. As the temperature is decreased, first a sharp transition into a highly pinning state (seen as dark blue) occurs between frames 2 and 3 in a temperature range of < 10 mK. Next, a lower-pinning phase (seen as yellow) appears at the top right corner of the frame (closest to the sample centre) and progresses gradually towards the sample edge. We note that the phase front separating the two phases is stable over the time period of frame acquisition. An unusual interaction of the phase boundary with sample inhomogeneity is seen in frame 7, where the weak-pinning bright phase meanders around the inhomogeneity. Phase coexistence is observed over a temperature range of ~ 50 mK.

several applied d.c. fields. Since a response of a larger part of the sample is now being taken into account, the width in temperature of the phase coexistence region in Fig. 4 also appears to be larger than in the previous imaging experiment done on a small area. It is thus the geometry that strongly affects the temperature range of phase coexistence detected in a given sample.

We now discuss the possible mechanisms and implications of the observed phenomena. The strong history dependence and the apparent stability of the two-phase coexistence region imply that thermal fluctuations are insufficient for the system to reach a unique state. Thus, it is the competition between the disorder and the elasticity of the vortex lines that is responsible for this transition. A higher critical current implies that the vortex lines adapt better to the sample disorder potential, and therefore the higher-pinning

vortex phase possesses a higher degree of topological disorder. Both the pinning strength and the elasticity are, however, temperature dependent, which is why the transformation is observed by varying the temperature. In general agreement with the edge contamination model⁹ of Paltiel *et al.*, the strong-pinning phase appears to nucleate near the sample edges and progressively enter the sample bulk as T_p is approached from below. We note that the pinning strength does not change continuously, but instead two distinct phases—with different j_c s—appear and coexist during the transformation. This is the hallmark of a first-order phase transition.

The important questions, therefore, are what drives the interface in the transition region and what stabilizes the particular phase mixture. In this context, the role of the applied a.c. field in driving the phase boundary was studied by an additional set of experiments.

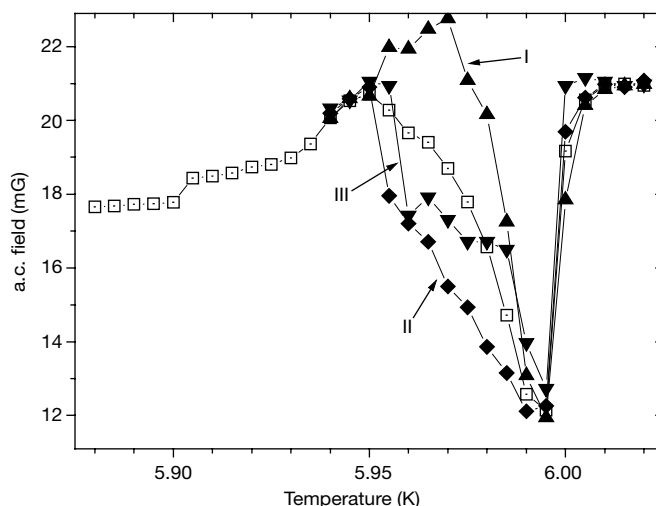


Figure 2 Plot of 'local' B_{ac} averaged over the areas I–III shown in frame 1 of Fig. 1b. Each area is $5 \mu\text{m}^2$; data from areas I, II and III are shown as up triangles, diamonds and down triangles, respectively. The position dependence of the peak-effect phenomenon is seen. The frame-averaged B_{ac} (squares) is also shown in the plot for comparison.

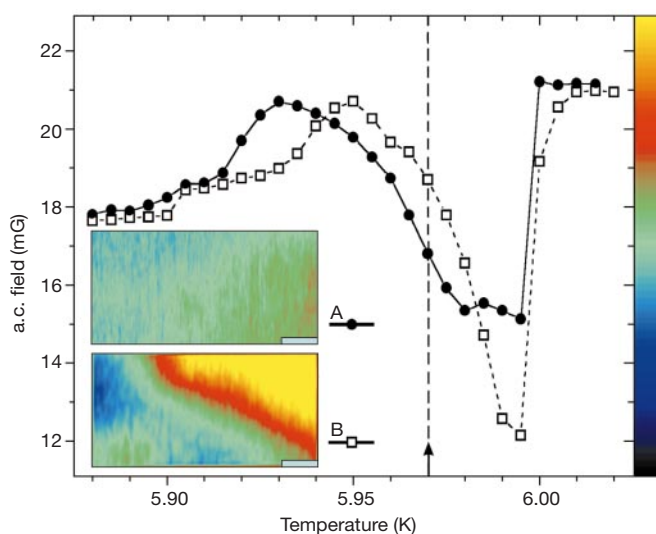


Figure 3 Plot of frame-averaged B_{ac} . Filled circles, data marked A in inset measured while the temperature was increased from 5.88 K to above T_c . The sample was first cooled in the ambient field of < 2 G to 5.7 K, and the field $B = 270$ G was applied at that temperature. The result of the field-cooling experiment is also shown (data marked B in inset, open squares) for comparison. A difference in peak amplitude and shape is observed. Inset: top, a representative frame for $T = 5.97$ K. The strong-pinning phase

appears from the left side of the frame, and gradually fills up the field of view. The interface between the phases is not as distinct as in the field-cooled case; also, it propagates along a different path and the j_c of the high-pinning phase is apparently lower. A corresponding frame from the field-cooling experiment is shown for comparison below.

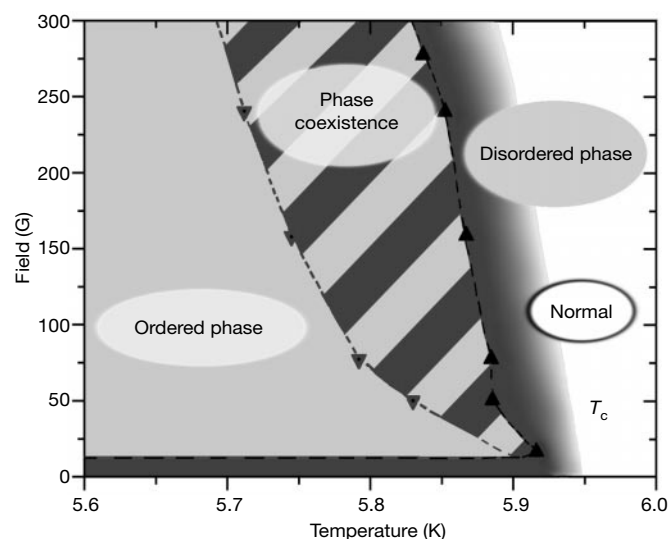


Figure 4 The 'phase diagram' for the peak effect. Phase coexistence occurs in a large temperature range below the superconducting transition, as sketched in the plot. The data points were measured on a different 2H-NbSe₂ sample (5 × 2 × 0.12 mm), with a static Hall sensor (see text) positioned ~10 μm above the sample surface. The peak effect is observed down to $B \approx 20$ G; for lower fields, a disordered phase persists presumably in the entire temperature range below T_c .

In a range of a.c. fields of 10–30 mG, the position and width of the peak do not depend on the a.c. field amplitude. However, the fine details of the local phase transformation do. Also, on several occasions, for a small a.c. field amplitude of ~5 mG, no peak effect was observed, suggesting that a supercooled disordered phase can be stabilized by the local disorder. In the opposite case of a large a.c. field amplitude of 40 mG, the magnitude of the peak effect decreases, suggesting that a superheated ordered phase can be stabilized by the drive.

Additionally, we speculate that the interface is driven by the change in the circulating shielding current that flows when the field or temperature is changed. The origin of the shielding current is in the equilibrium diamagnetism of the superconductor, including a possible step change in it due to a first-order phase transition^{10,11}. A jump in local induction¹¹ associated with the first-order phase transition could potentially create a current density at least an order of magnitude larger than the shielding a.c. current. The flow of current is then self-balanced between the two phases, and its confinement within the sample geometry determines the overall spatial phase distribution. In a similar fashion to the first-order melting transition in high-temperature superconductors¹², small variations in the pinning strength across the sample may lead to the observed complex shape of the phase boundary.

An important implication of our work is that the two-phase nature of the peak-effect region should be taken into account when producing a coherent explanation of the numerous experimental results. In fact, a consideration of redistribution of current in a two-phase system leads naturally to a simple quantitative model (M.M. *et al.*, manuscript in preparation) of numerous experimentally observed transport anomalies. The two-phase nature of the transition region, as opposed to the commonly considered spatially homogeneous character, is crucial for a proper analysis of the structural data¹³. Also, our result may be relevant to the explanation of the "second peak" phenomenon¹⁴ observed in the high-temperature superconductors.

The peak effect thus appears to be a rare experimental realization of a disorder-driven non-thermal phase transformation in a randomly disordered system. This class of phase transitions is of considerable

present interest, and we hope our results will lead to further theoretical and experimental explorations of the subject. □

Received 29 September; accepted 27 November 2000.

1. Pippard, A. B. A possible mechanism for the peak effect in type II superconductors. *Phil. Mag.* **19**, 217–220 (1969).
2. Wordenweber, R., Kes, P. H. & Tsuei, C. C. Peak and history effect in two dimensional collective flux pinning. *Phys. Rev. B* **33**, 3172–3180 (1986).
3. Bhattacharya, S. & Higgins, M. J. Dynamics of a disordered flux line lattice. *Phys. Rev. Lett.* **70**, 2617–2620 (1993).
4. Henderson, W. *et al.* Metastability and glassy behavior of a driven flux line lattice. *Phys. Rev. Lett.* **77**, 2077–2080 (1996).
5. Henderson, W., Andrei, E. Y. & Higgins, M. J. Plastic motion of a vortex lattice driven by alternating current. *Phys. Rev. Lett.* **81**, 2352–2355 (1998).
6. Banerjee, S. S. *et al.* Disorder, metastability, and history dependence in transformations of a vortex lattice. *Phys. Rev. B* **59**, 6043–6046 (1999).
7. Chang, A. M. *et al.* Scanning Hall probe microscopy. *Appl. Phys. Lett.* **61**, 1974–1976 (1992).
8. Gilchrist, J. & Konczykowski, M. Superconductor screen viewed as one or two inductive loops. *Physica C* **212**, 43–60 (1993).
9. Paltiel, Y. *et al.* Dynamic instabilities and memory effects in vortex matter. *Nature* **403**, 398–401 (2000).
10. Zeldov, E. *et al.* Thermodynamic observation of first-order vortex lattice melting transition. *Nature* **375**, 373–376 (1995).
11. Ravikumar, G. *et al.* Step change in equilibrium magnetization across the peak effect in 2H-NbSe₂. *Physica C* **322**, 145–150 (1999).
12. Soibel, A. *et al.* Imaging the vortex-lattice melting process in the presence of disorder. *Nature* **406**, 282–287 (2000).
13. Gammel, P. L. *et al.* Structure and correlation of a flux line lattice in crystalline Nb through peak effect. *Phys. Rev. Lett.* **80**, 833–836 (1998).
14. Khaykovich, B. *et al.* Vortex-lattice phase transitions in Bi₂Sr₂CaCu₂O₈ crystals with different oxygen stoichiometry. *Phys. Rev. Lett.* **76**, 2555–2558 (1996).

Acknowledgements

We thank E. Zeldov and E. Andrei for helpful discussions.

Correspondence and requests for materials should be addressed to M.M. (e-mail: marchev@research.nj.nec.com).

An effective gravitational temperature for sedimentation

P. N. Segrè*, F. Liu†, P. Umbanhowar‡ & D. A. Weitz*

* Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

† Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

‡ Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA

The slow sedimentation of suspensions of solid particles in a fluid results in complex phenomena that are poorly understood. For a low volume fraction (ϕ) of particles, long-range hydrodynamic interactions result in surprising spatial correlations¹ in the velocity fluctuations; these are reminiscent of turbulence, even though the Reynolds number is very low^{2–4}. At higher values of ϕ , the behaviour of sedimentation remains unclear; the upward back-flow of fluid becomes increasingly important, while collisions and crowding further complicate inter-particle interactions^{5–8}. Concepts from equilibrium statistical mechanics could in principle be used to describe the fluctuations and thereby provide a unified picture of sedimentation, but one essential ingredient—an effective temperature that provides a mechanism for thermalization—is missing. Here we show that the gravitational energy of fluctuations in particle number can act as an effective temperature. Moreover, we demonstrate that the high- ϕ behaviour is in fact identical to that at low ϕ , provided that the suspension viscosity and sedimentation velocity are scaled appropriately, and that the effects of particle packing are included.

We studied the sedimentation of relatively high concentrations ($0.05 \leq \phi \leq 0.50$) of nearly monodisperse spherical glass beads of mean radii $a = 50 \pm 5 \mu\text{m}$ dispersed in solutions of water and glycerol (viscosity, $\eta \approx 20\text{--}70 \text{ cP}$) that nearly match the refractive index of the particles. Thermally induced brownian motion is negligible compared to gravitational settling, with the Peclet number being $\text{Pe} \approx 10^7$; particle inertia is also negligible, with the particle Reynolds number being $\text{Re} \approx 10^{-4}$. The sample cells were typically 6 cm wide, 0.3 cm deep and 30 cm high, and were placed in a stirred water bath at temperature $T = 23 \pm 1^\circ\text{C}$. The cell was illuminated from behind, and images were collected from the median plane with a CCD (charge-coupled device) camera, using a depth of field of $\sim 0.5\text{--}1 \text{ mm}$. These images were digitized, and velocity profiles were calculated using standard particle-image-velocimetry (PIV) techniques⁹. This yielded two-dimensional velocity maps of 29×39 vectors, each of which is the local average of the velocities of two to four spheres. Initially random dispersions were prepared by vigorously shaking the cell, and measurements were started after the sedimentation front had settled about one-fifth of the cell height, allowing any initial transients to decay. Images were collected from regions of the cell that were far from both the sedimentation front at the top, and the concentrated sediment at the bottom of the cell.

We subtract the mean velocity to focus on velocity fluctuations $\delta\mathbf{V} = \mathbf{V} - V(\phi)\hat{z}$, where $V(\phi)\hat{z}$ is the average settling velocity due to gravity; in Fig. 1 we show typical spatial maps of $\delta\mathbf{V}$ for a sample with $\phi = 0.30$. The magnitudes of the fluctuations are very large, of the order of the average velocity. Moreover, similar to the behaviour at much lower ϕ ¹, the velocity fluctuations are clearly strongly correlated spatially over length scales that are much greater than a . These spatially correlated patterns evolve over time, as illustrated by the sequence of $\delta\mathbf{V}$ maps shown in Fig. 1; it is this complex spatial and temporal pattern which is reminiscent of turbulence (P. M. Chaikin, personal communication and ref 3). Several prominent correlated regions seen in Fig. 1a remain clearly identifiable in Fig. 1b, which is after $\Delta t = 18 \text{ s}$, corresponding to an average settling of $\sim 10a$; however, after $\Delta t = 36 \text{ s}$, corresponding to an average settling of $\sim 20a$, the patterns are nearly completely decorrelated, as shown in Fig. 1c. By examining many such images, we determine the root mean square (r.m.s.) magnitude of the velocity fluctuations, both parallel, $\Delta V_{\parallel}$, and perpendicular, $\Delta V_{\perp}$, to gravity, and plot these as a function of ϕ in Fig. 2. The magnitudes of the r.m.s. fluctuations are comparable to the mean sedimentation velocities, $\Delta V/V(\phi) \approx 1$; thus mixing is substantial during settling. Moreover, $\Delta V/V(\phi)$ is approximately constant for the volume fractions used in this work, $0.05 \leq \phi \leq 0.50$; by contrast, earlier results¹ in very dilute suspensions found a rapid increase for $10^{-4} \leq \phi \leq 0.03$.

The spatial extent of the velocity correlations is quantified by calculating $C_z(z) = \langle \delta V_z(0)\delta V_z(z) \rangle / \Delta V_{\parallel}^2$, where z is parallel to gravity. These correlation functions decay exponentially, as shown in Fig. 3 inset, providing a measure of the characteristic length scale of the correlations of the velocity fluctuations, $\xi_{\parallel}$. We similarly determine the transverse correlation length in the direction perpendicular to gravity, $\xi_{\perp}$, for motion parallel to gravity (found from $C_z(x)$, where x is perpendicular to gravity). We plot the ϕ -dependence of both correlation lengths in Fig. 3; both agree well with extrapolations of the behaviour measured for very dilute suspensions, $(\phi \leq 0.03)$ ¹, $\xi_{\perp} \approx 7a\phi^{-1/3}$ and $\xi_{\parallel} \approx 11a\phi^{-1/3}$, as shown by the dashed and dotted lines, respectively. This agreement is surprising because of the significantly different mechanisms influencing particle interactions as ϕ increases; whereas hydrodynamic interactions play a crucial role at all ϕ , short-range lubrication forces, solvent back-flow and excluded-volume effects become increasingly important with increasing ϕ .

We also determine the characteristic timescale over which the velocity fluctuations remain correlated by measuring their temporal correlation at a fixed spatial location that is stationary in the frame of the mean sedimentation velocity,

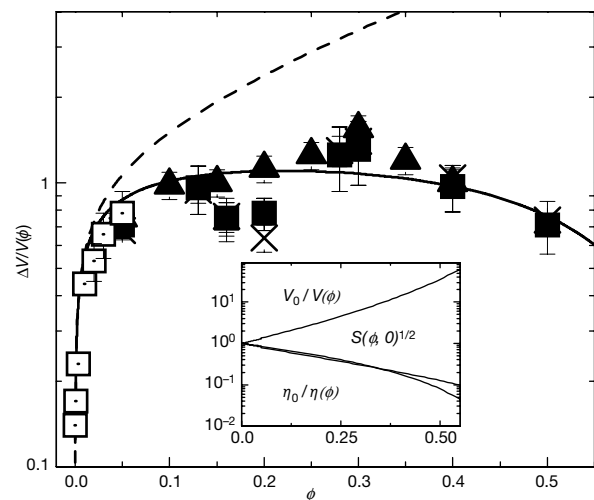


Figure 2 Normalized velocity fluctuations $\Delta V/V(\phi)$ as a function of ϕ . These fluctuations are in the directions parallel ($\parallel$; filled squares, this work; triangles, ref. 12; open squares, ref. 1) and perpendicular ($\perp$; multiplied by 2, crosses, this work) to gravity. The solid and dashed lines are from equation (2), with $C_0 = 2.6$, with and without $(S(\phi, 0) = 1)$ excluded volume corrections. Inset: ϕ -dependencies of sedimentation velocity, viscosity, and structure factors used in equation (2).

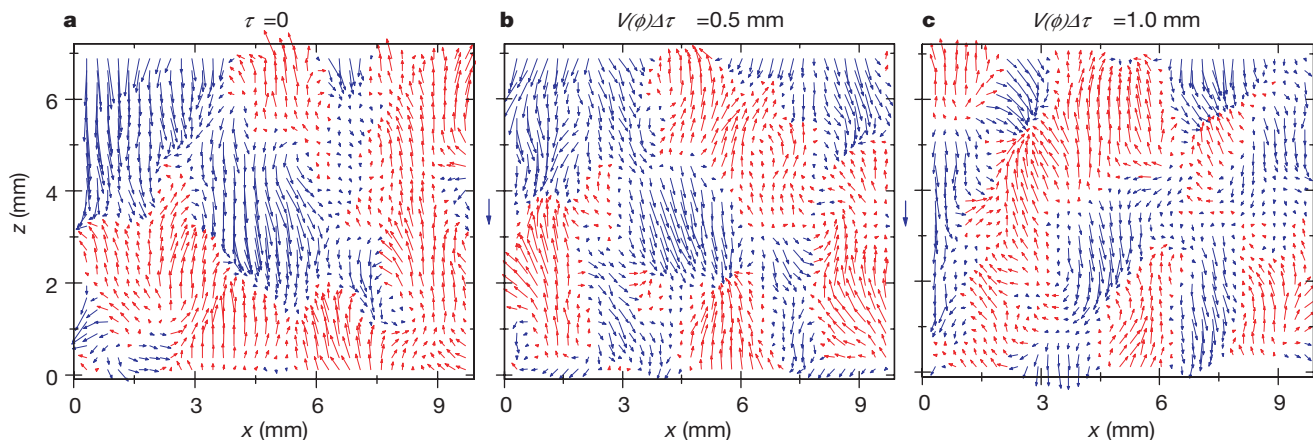


Figure 1 PIV velocity fluctuation ($\delta\mathbf{V} = \mathbf{V} - V(\phi)\hat{z}$) vector maps of a typical sample at $\phi = 0.30$. Data are shown at relative times $t = 0$ (a), $\Delta t = 18 \text{ s}$ (b) and $\Delta t = 36 \text{ s}$ (c). Blue and red colouring is used to distinguish the up and down directions, respectively. The

mean sedimentation velocity, $V(\phi) = 0.028 \text{ mm s}^{-1}$, is represented to scale by the single arrows next to each frame. Each vector represents the local instantaneous velocity of 2–4 particles.

$C(\tau) = \langle \delta V_z(0) \delta V_z(\tau) \rangle / \Delta V^2$. These correlation functions decay exponentially, with a characteristic decay time, τ_d , as shown in Fig. 4 inset. This decay time increases with increasing ϕ , and is comparable to the transit time for a velocity fluctuation, $\tau_\xi = \xi / \Delta V$. Thus, the regions of correlated velocity fluctuations persist long enough that particles within these regions can traverse the correlation length; similar behaviour was observed using ultrasonic scattering from fluidized beds at high ϕ (ref. 10). This suggests a simple physical picture: particles are advected within a correlated velocity fluctuation, travelling a distance of $\sim \xi$ until the correlations decay in a time, τ_d , whereupon the particle becomes entrained in a new region of correlated fluctuating velocity. This behaviour is analogous to a diffusion process, with the mean free path of the diffusion determined by ξ and the mean free time by τ_d . To test this analogy, we compare the diffusion coefficients determined by $D(\phi) = \Delta V^2 \tau_d$ with the self-diffusion coefficients $D_s(\phi)$ measured by tracking tagged particles^{11,12}; as shown in Fig. 4, they are in very good agreement. This establishes a direct relationship between the collective correlations in the velocity fluctuations and the self-diffusion of individual particles.

A simple energy-balance model can be used to relate the magnitude of the velocity fluctuations ΔV to the correlation lengths ξ (refs 1, 6). We assume that the correlations in the velocity fluctuations result from a fluctuation in particle density, of spatial extent ξ , corresponding to a fluctuation in particle mass of Δm_ξ . We determine the magnitude of this fluctuation by balancing the gravitational potential energy gained as this fluctuation moves up or down relative to the background suspension with the energy lost to viscous dissipation due to shear. Thus, for a system in steady state we equate the rate of energy gain, $dE_g/dt = \Delta m_\xi g \Delta V$, with the rate of energy dissipation, $dE_v/dt = (6\pi\eta(\phi)\xi\Delta V)\Delta V$, where $\eta(\phi)$ is the ϕ -dependent viscosity of the suspension. This leads to:

$$\Delta V = \frac{\Delta m_\xi g}{6\pi\eta(\phi)\xi} \quad (1)$$

We assume that mass fluctuations arise from random fluctuations in particle number in a volume of ξ^3 , $\Delta m_\xi = \Delta N_\xi m_1$, where $m_1 = (4\pi/3)a^3\Delta\rho$ is the buoyant mass of a single particle of volume v_p whose density differs from that of the fluid by $\Delta\rho$. The average number of particles in this volume is $\langle N_\xi \rangle = (\xi/a)^3\phi$, which is independent of ϕ , given the measured scaling of ξ . We expect ΔN_ξ

to scale as $\langle N_\xi \rangle^{1/2}$; however, we must also incorporate the effects of excluded volume as ϕ increases, and thus obtain $\Delta N_\xi = [\langle N_\xi \rangle S(\phi, 0)]^{1/2}$, where $S(\phi, 0)$ is the static structure factor in the small wavevector limit¹³. Thus, for the parallel velocity fluctuations, we obtain

$$\frac{\Delta V}{V(\phi)} = C_0 \phi^{1/3} \left[\frac{V_0}{V(\phi)} \frac{\eta_0}{\eta(\phi)} \sqrt{S(\phi, 0)} \right] \quad (2)$$

where $C_0 = \sqrt{11} \approx 3.3$ and $V_0 = (2/9)\Delta\rho g a^2/\eta_0$ is the Stokes settling velocity at infinite dilution in a fluid of viscosity η_0 .

To compare this prediction with our data, we use the measured¹⁴ functional form for the sedimentation velocity, $V(\phi)/V_0 = (1 - \phi)^5$, which is in good agreement with our data. For the viscosity and structure factor, we use expressions for suspensions of brownian hard spheres in thermal equilibrium. We obtain the static structure factor from the isothermal compressibility, $S(\phi, 0) = (k_B T)/\nu_p = [\partial \Pi / \partial \phi]^{-1}$, where we use the Carnahan–Starling equation of state for the osmotic pressure¹³, $\Pi(\phi) = (k_B T/\nu_p)[(\phi + \phi^2 + \phi^3 - \phi^4)/(1 - \phi)^3]$. Since the shear rates generated by the fluctuations, $\dot{\gamma} = \Delta V/\xi$, are much larger than the rate of thermal brownian diffusion, D_B/a^2 , ($Pe = \dot{\gamma}/(D_B/a^2) \approx 10^4$), we obtain the viscosity from the functional form measured in the high-shear-rate, or high-Peclet-number, limit, where brownian effects are negligible and configurational effects dominate^{15,16}, $\eta(\phi)/\eta_0 = (1 - \phi/0.71)^{-2}$. Using the functional forms in equation (2), an excellent agreement is obtained with the measured value of the velocity fluctuations as a function of ϕ , as shown by the solid line in Fig. 2. This agreement is notable especially given the strong ϕ -dependencies of $V(\phi)$, $\eta(\phi)$ and $S(\phi, 0)$, shown in Fig. 2 inset; we note in particular that the effects of volume exclusion are critically important. We emphasize that we use the functional forms for a thermal system to describe an athermal system. However, the main effect of thermalization is to ensure that the system explores phase space; although the thermalization mechanism is different for a sedimenting system as compared to a brownian system, sedimenting particles are still thermalized in that they can explore accessible phase space. Thus the effects of volume exclusion at high ϕ are similar in both cases, allowing us to use these expressions.

As a further test of the consequences of the thermalization, we also calculate the behaviour of the self-diffusion coefficient using

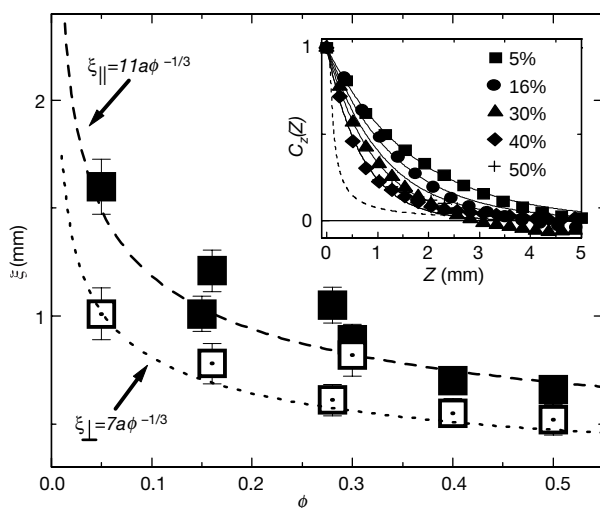


Figure 3 Correlation lengths, ξ , as a function of ϕ . Filled squares, longitudinal correlation lengths, $\xi_{||}$; open squares, transverse correlation lengths, $\xi_{\perp}$. The dashed lines are extrapolations of the scaling results found at very low ϕ (ref. 1). Inset: longitudinal velocity correlation functions $C_v(z)$ as a function of z , the distance in the direction of gravity. The dashed line is the theoretical result for an isolated sphere of size a , with no cooperative effects¹³.

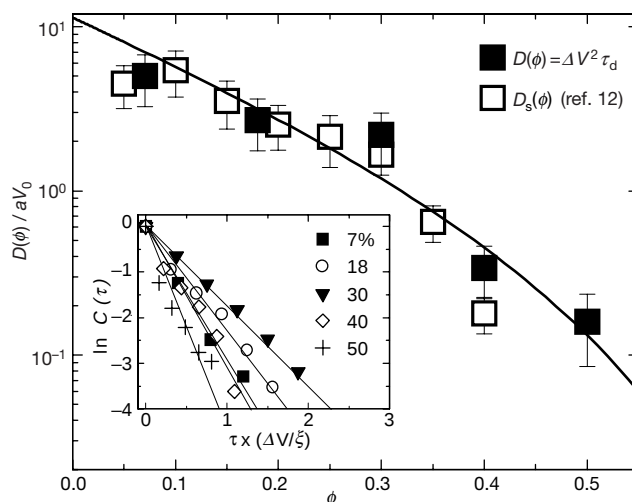


Figure 4 Normalized diffusion coefficients $D(\phi) = \Delta V^2 \tau_d$ from this work and $D_s(\phi)$ from ref. 12. Solid line: $D(\phi) = 11.4aV_0\{\eta_0/[\eta(\phi)]\}[S(\phi, 0)]^{1/2}$. Inset: time correlation functions, $C(\tau)$, of the velocity vector maps plotted as a function of τ normalized by the ϕ -dependent transit time for a velocity fluctuation $\xi/\Delta V$. The decorrelation time τ_d is found from fits to $C(\tau) = \exp(-\tau/\tau_d)$, shown as solid lines, yielding on average $\tau_d \approx \xi/2.5\Delta V$.

$D(\phi) = \Delta V^2 \tau_d$. We approximate the behaviour of the decay time with $\tau_d = \xi/2.5\Delta V$, and use the measured functional form for ξ , and the expression for ΔV in equation (2), thereby obtaining a new analytic expression for the diffusion coefficient, $D(\phi) = 11.4aV_0\{\eta_0/[\eta(\phi)]\}[S(\phi, 0)]^{1/2}$. We again obtain very good agreement with the data, as shown by the solid line in Fig. 4. This agreement further validates our use of thermal equilibrium quantities to describe this highly athermal system.

We can obtain insight into the origin of the thermalization of the sedimenting particles by considering the diffusion coefficient. We take the approximate relation based on our data, $D(\phi) \approx \Delta V\xi/2.5$, and use equation (1) for the velocity fluctuations, to obtain

$$D(\phi) = \frac{C_1 \Delta m_g g \xi}{6\pi\eta(\phi)\xi} \quad (3)$$

where $C_1 \approx 0.4$. This expression has the same functional form as the Stokes–Einstein equation for the diffusion coefficient of thermalized particles. This allows us to identify the effective temperature, $k_B T = C_1 \Delta m_g g \xi$, which provides the equivalent of thermal energy to the system allowing it to explore phase space, and resulting in the diffusive motion of the particles. But, here it is clearly the gravitational energy of the fluctuation in particle density which drives the thermalization. We note that it is only the fluctuating part of the energy, that arising from Δm_g , which drives the system and causes the diffusive motion. By contrast, the full buoyant mass of the particles is much larger, but contributes only to the average settling, and not to the fluctuations which are the origin of the thermalization. The conversion of potential energy into viscous shear energy, rather than translational kinetic energy, is a consequence of the vanishingly small Reynolds number; Re is related to the ratio of kinetic to shear energy.

Finally, we note that the effective temperature can only be defined for the parallel component of the motion; in the perpendicular direction, there is no gravitational potential energy. The difference in the parallel and perpendicular components of the diffusion coefficients suggest that the effective temperature also exhibits an asymmetry¹²; for the perpendicular temperature it must be the coupling of the two components of the velocity fluctuations that causes the thermalization and leads to the diffusive motion. Nevertheless, these results provide a clear measure of the effective temperature for sedimenting particles, and offer new insight into the relationship between the diffusive motion of individual particles and collective correlations of the velocity fluctuations. □

Received 6 October; accepted 30 November 2000.

1. Segrè, P. N., Herbolzheimer, E. & Chaikin, P. M. Long-range correlations in sedimentation. *Phys. Rev. Lett.* **79**, 2574–2577 (1997).
2. Xue, J.-Z., Herbolzheimer, E., Rutgers, M. A., Russel, W. B. & Chaikin, P. M. Diffusion, dispersion, and settling of hard spheres. *Phys. Rev. Lett.* **69**, 1715–1718 (1992).
3. Tong, T. & Ackerson, B. J. Analogies between colloidal sedimentation and turbulent convection at high Prandtl numbers. *Phys. Rev. E* **58**, R6931–R6934 (1998).
4. Levine, A., Ramaswamy, S., Frey, E. & Bruinsma, R. Screened and unscreened phases in sedimenting suspensions. *Phys. Rev. Lett.* **81**, 5944–5947 (1998).
5. Caflisch, R. E. & Luke, J. H. C. Variance in the sedimentation speed of a suspension. *Phys. Fluids* **28**, 759–760 (1985).
6. Hinch, E. J. in *Disorder and Mixing* (eds Guyon, E., Nadal, J.-P. et al. Pomeau, Y.) 153–185 (Kluwer Academic, Dordrecht, 1988).
7. Koch, D. L. & Shaqfeh, E. S. G. Screening in sedimenting suspensions. *J. Fluid Mech.* **224**, 276–303 (1991).
8. Brenner, M. P. Screening mechanisms in sedimentation. *Phys. Fluids* **11**, 754–772 (1999).
9. Adrian, R. J. Particle-imaging techniques for experimental fluid mechanics. *Annu. Rev. Fluid Mech.* **23**, 261–304 (1991).
10. Cowan, M. L., Page, J. H. & Weitz, D. A. Velocity correlations in fluidized suspensions probed by ultrasonic correlation spectroscopy. *Phys. Rev. Lett.* **85**, 453–456 (2000).
11. Nicolai, H. & Guazzelli, E. Effect of the vessel size on the hydrodynamic diffusion of sedimenting spheres. *Phys. Fluids* **7**, 3–5 (1995).
12. Nicolai, H., Herzhaft, B., Hinch, E. J., Oger, L. & Guazzelli, E. Particle velocity fluctuations and hydrodynamic self-diffusion of sedimenting non-Brownian spheres. *Phys. Fluids* **7**, 12–23 (1995).
13. Russel, W. B., Saville, D. A. & Schowalter, W. R. *Colloidal Dispersions* (Cambridge Univ. Press, Cambridge, 1989).

14. Richardson, J. F. & Zaki, W. N. Sedimentation and fluidization I. *Trans. Inst. Chem. Eng.* **32**, 35–53 (1954).
15. de Kruif, C. G., van Iersel, E. M. F., Vrij, A. & Russel, W. B. Hard sphere colloidal dispersions: viscosity as a function of shear rate and volume fraction. *J. Phys. Chem.* **83**, 4717–4725 (1985).
16. Bender, J. W. & Wagner, N. J. Reversible shear thickening in monodisperse and bidisperse colloidal dispersions. *J. Rheol.* **40**, 899–916 (1996).

Acknowledgements

We thank M. Brenner, S. Tee, P. Tong, A. J. C. Ladd, B. J. Ackerson, P. Mucha and A. Levine for discussions. This work was supported by NASA, NSF and the donors of the Petroleum Research Fund, administered by the ACS. Current address of P.N.S. is NASA MSFC, Huntsville, AL 35802 (Phil.Segre@msfc.nasa.gov).

Correspondence and requests for materials should be addressed to D.A.W. (e-mail: weitz@deas.harvard.edu) or P.N.S.

Pumping of nutrients to ocean surface waters by the action of propagating planetary waves

B. Mete Uz, James A. Yoder & Vladimir Osychyn

Graduate School of Oceanography, University of Rhode Island, S. Ferry Road, Narragansett, Rhode Island 02882, USA

Primary productivity in the oceans is limited by the lack of nutrients in surface waters. These nutrients are mostly supplied from nutrient-rich subsurface waters through upwelling and vertical mixing¹, but in the ocean gyres these mechanisms do not fully account for the observed productivity². Recently, the upward pumping of nutrients, through the action of eddies, has been shown to account for the remainder of the primary productivity; however, these were regional studies which focused on mesoscale (100-km-scale) eddies^{3–6}. Here we analyse remotely sensed chlorophyll and sea-surface-height data collected over two years and show that 1,000-km-scale planetary waves, which propagate in a westward direction in the oceans, are associated with about 5 to 20% of the observed variability in chlorophyll concentration (after low-frequency and large-scale variations are removed from the data). Enhanced primary production is the likely explanation for this observation, and if that is the case, propagating disturbances introduce nutrients to surface waters on a global scale—similar to the nutrient pumping that occurs within distinct eddies.

Much of the motion in the open ocean is baroclinic: at the surface, geostrophic flow balances the pressure gradient caused by a sloping sea surface. The geostrophic current is typically strongest near the surface, and decreases with depth. Because the horizontal pressure gradient must similarly decrease with depth in order to maintain geostrophic balance, constant-density surfaces must slope in the opposite direction to the sea surface. As a result, a baroclinic disturbance that displaces the constant-density surfaces upwards has a negative sea surface height (SSH) anomaly, which can be measured by altimeters. These disturbances propagate westward—as planetary waves or eddies—mostly due to the variation of the Coriolis parameter with latitude. Westward-propagating first-mode baroclinic features dominate the propagating components of the SSH satellite-sensed record from the TOPEX/POSEIDON mission^{7–9}. They have also been observed in images of sea surface temperature from the satellite-borne Advanced Very High Resolution Radiometer^{10–12}.

If the displacement of constant-density surfaces by baroclinic disturbances brought new nutrients into the euphotic zone, the concentration of phytoplankton chlorophyll would be expected to

increase above the disturbance. As the disturbance propagates westward, there would be enhanced production near the leading edge and fall-out and grazing of the biomass anomaly near the trailing edge, so that the chlorophyll anomaly would also propagate with the disturbance. If this were to happen, chlorophyll variability would have a component that was coherent with the SSH variations.

To look for this signature in the satellite imagery, we sought large-scale anomalies in SeaWiFS¹³ remotely sensed oceanic chlorophyll concentration that propagate westward in coherence with SSH anomalies. Even though the eddy pumping mechanism was originally intended for mesoscale eddies, our study focuses on the larger scales of Rossby waves; this larger scale enables us to do more spatial smoothing and compositing (replacing cloud-masked pixels of one image with corresponding pixels of a consecutive image) over time. The compression of the upper layers of the ocean by a Rossby wave is very small dynamically to the compression caused by a small-amplitude eddy, and the same mechanism of nutrient enhancement should apply. With a nonlinear eddy that transports water, the pumping would occur only once—during the spin-up of the eddy.

Propagating anomalies were readily observed as slanted features in time–longitude diagrams (Fig. 1) once data were bandpass-filtered to remove noise and variability at small scales as well as

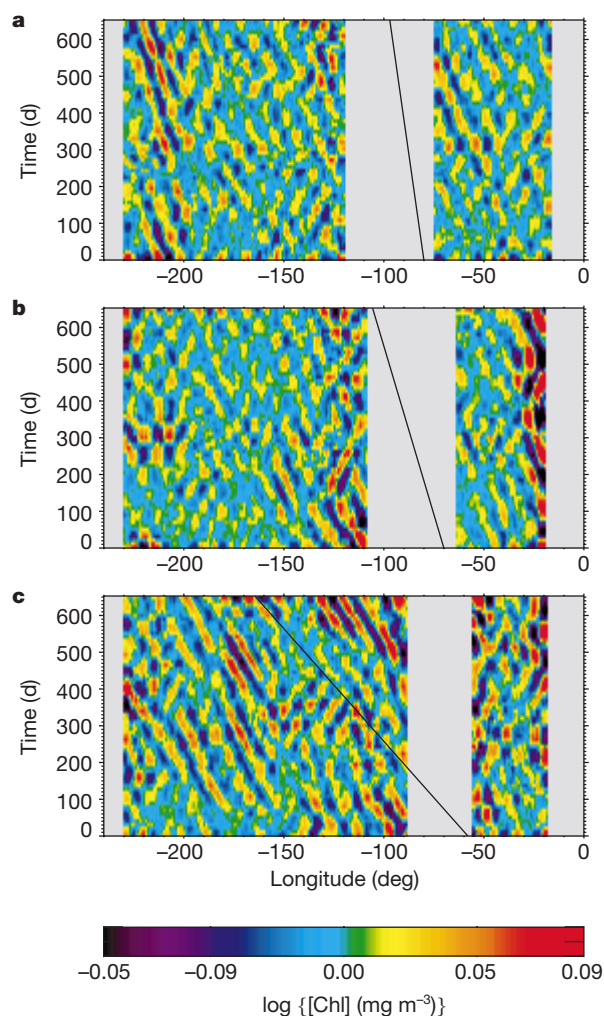


Figure 1 Time–longitude diagrams of the bandpass-filtered SeaWiFS chlorophyll concentration data. The figure shows zonal propagation of anomalies as slanted features. The phase speeds of linear non-dispersive first-mode baroclinic Rossby waves are shown by a solid line at each latitude (**a**, 30° N; **b**, 20° N; **c**, 10° N). The variation with latitude of the propagation speed of chlorophyll anomalies may be seen as an animation sequence at http://po.gso.uri.edu/~uz/chl_anomaly.htm.

seasonal and other very-large-scale variations. These anomalies persisted for a long time. For example, a high-chlorophyll-concentration anomaly at 30° N which was at longitude –200° (160° E) around day 150 (150 days after 7 September 1997) was still visible on day 550, by which time it had moved to longitude –225° (135° E).

We used spectral analysis to investigate the coherence between chlorophyll and SSH features in order to separate different scales of variability and propagation directions. We calculated the power spectral density of—and coherence between—chlorophyll and SSH for individual ocean basins. At most latitudes, westward-propagating features made up almost all of the spectral density of SSH, whereas chlorophyll spectral density showed more spread. Part of the spread in the chlorophyll spectral density may be due to non-propagating phenomena such as the expansion of blooms in both directions. High coherence between chlorophyll and SSH was mostly in the westward-propagation quadrants (Fig. 2).

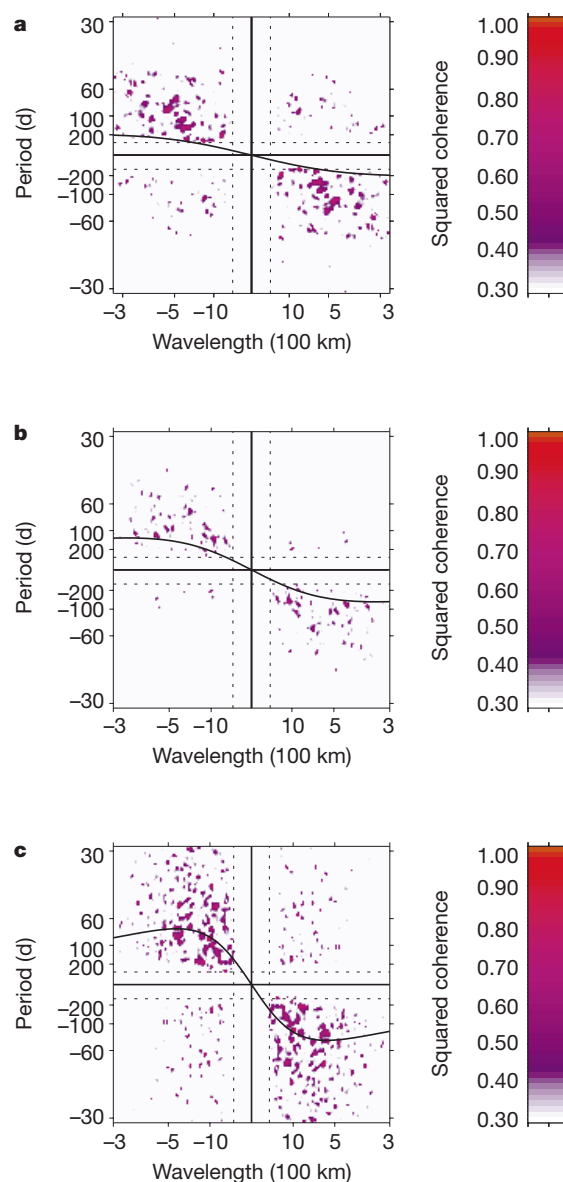


Figure 2 Wavenumber–frequency spectra of coherence between log(chlorophyll concentration) and sea surface height. Most of the spectral density is in the second and fourth quadrants, where propagation is westward. Data shown are from the Pacific Ocean at latitudes of 30° N (**a**), 20° N (**b**) and 10° N (**c**). Dotted lines bracketing the axes show cut-off points for the highpass filters. The linear dispersion relation is shown by a solid line.

Phase speeds of planetary waves measured with the TOPEX/POSEIDON altimeter have been seen to vary between 0.5 and 3 times the phase speed from the linear dispersion⁷, though a recent study reports that with different methods the linear phase speed is within the error bars¹⁴. Nevertheless, the theoretical phase speed of baroclinic planetary waves may not be adequate for a precise comparison with measurements. Qualitatively, the propagation speeds of chlorophyll anomalies that we saw in the time–longitude diagrams were similar to the theoretical propagation speed of first-mode baroclinic planetary waves, and (more importantly) they varied with latitude in the same way.

We calculated propagation speeds from the chlorophyll and chlorophyll–SSH coherence spectra, and compared these to the non-dispersive linear phase speed calculated using climatological density profiles¹⁵ (Fig. 3). Between 10° and 25° in both hemispheres, estimates from chlorophyll, SSH and chlorophyll–SSH coherence were all similar and within a factor of two of the theoretical phase speed. At higher latitudes (>25°), propagation speeds from chlorophyll spectra were significantly higher than the rest. This deviation must be caused by phenomena not related to SSH, because the propagation speed of the SSH coherent features (the estimate from chlorophyll–SSH coherence) remained close to the phase speed from SSH and the theoretical phase speed.

The confinement of high chlorophyll–SSH coherence in a localized area of the wavenumber–frequency domain allowed us to construct a filter to isolate chlorophyll features which were associated with the SSH anomalies. The variance of the coherent, westward-propagating signal was around 5–20% of the variance of the input (Fig. 4). These fractions depend on the choice of the threshold coherence, and are provided here as only a rough scale for the magnitude of the effect. This fraction varied with latitude, and between basins at any given latitude. There were local minima at latitudes corresponding to the middle of the subtropical gyres. The coherence estimate is sensitive to the relative magnitude of error in the data. If each data series is modelled as the sum of an ideal signal and random noise, wherever the amplitudes of the signals are reduced, the coherence estimate also goes down even though the coherence between the two ideal signals remains the same. This effect may be what we are seeing in the middle of the gyres, where the mean as well as the variability of chlorophyll is low and therefore the noise figure is high. The local minima at these locations may then be mostly in the estimate, and might not occur in the estimates

if more accurate data were available. Nevertheless, in all basins there were large areas where coherence was observed between chlorophyll and SSH. We note that our analysis used the entire time record; for chlorophyll and SSH features to be coherent, they must not only coincide in space at one point in time, but they must propagate together. A phase offset may exist between chlorophyll and SSH features without reducing the coherence as long as the offset remains constant.

The mechanism of enhanced primary production proposed in ref. 3 may explain how the chlorophyll anomaly can follow the baroclinic disturbance detected by the SSH anomaly. In this mechanism, the disturbance introduces nutrients into the euphotic zone; primary production is therefore enhanced, and eventually chlorophyll concentration increases in the surface water over the nutrient-enriched area. The nutrient enhancement may occur through the compression of near-surface layers, so that deeper layers richer in nutrients extend into the euphotic zone¹⁶. Alternatively, it may occur through isopycnal mixing within a tilted constant-density layer that protrudes into the euphotic layer over the disturbance¹⁷. There may also be a physiological delay between the introduction of the nutrients and the enhancement in chlorophyll concentration. For large-scale disturbances that have periods on the order of months, the phase offset caused by such delays should be very small, and unlikely to change much even as the anomaly propagates through areas of different planktonic community structures. The chlorophyll anomaly induced through this mechanism by a SSH anomaly should be easily detected with our methods.

The coherence between SSH and chlorophyll that we observe is consistent with the eddy pumping mechanism, but does not yet prove it: there are means other than enhanced primary production by which the coherence may be created. One of these is the advection of a horizontal background chlorophyll gradient by the geostrophic velocity associated with the baroclinic anomaly. In this case, chlorophyll acts as a passive tracer and the contrast visible in the imagery will be determined by the magnitude of the gradient and its displacement by the disturbance. Another possible mechanism is the shoaling of a deep chlorophyll maximum by the disturbance from an original depth too deep to be seen by the sensor. In the interior of the gyres, where the horizontal chlorophyll gradients are weak and the deep chlorophyll maximum occurs at greater depths, neither of these alternative mechanisms would be very effective. The presence of propagating chlorophyll anomalies in these areas supports the hypothesis of enhanced primary productivity. □

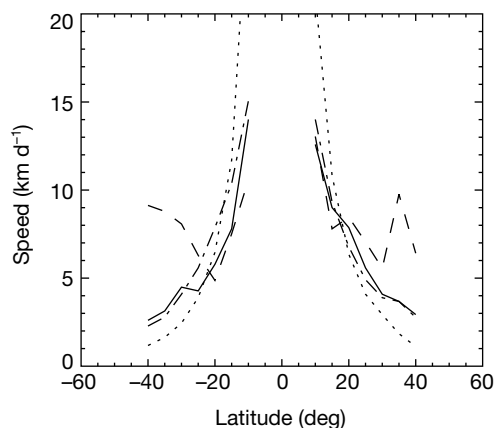


Figure 3 Propagation speeds of Chl and SSH anomalies. The figure shows that propagation speeds of chlorophyll anomalies associated with baroclinic features vary with latitude in nearly the same way as do the theoretical phase speeds of Rossby waves (dotted lines). The propagation speeds are estimated from the wavenumber spectra of chlorophyll (dash-dot line), sea surface height (SSH; dashed line) and chlorophyll–SSH coherence (solid lines) in the Pacific Ocean. The theoretical speeds are based on baroclinic radii of deformation from ref. 15.

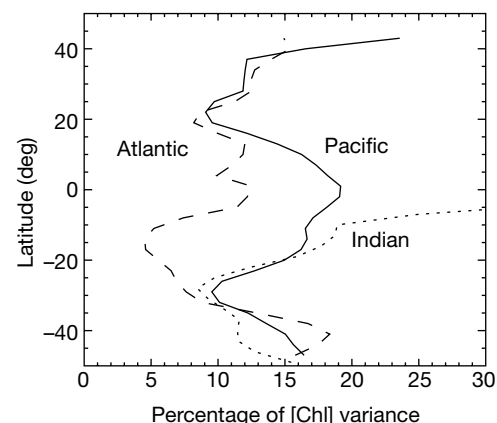


Figure 4 Features associated with westward-propagating baroclinic disturbances make up ~10% of the variance of the log(chlorophyll concentration) data. These chlorophyll data were filtered so as to eliminate variability at length scales larger than 25° and timescales longer than 260 d.

Methods

We acquired individual global area coverage swaths of SeaWiFS chlorophyll concentration from September 1997 to the end of December 1999, and mapped them to a cylindrical grid at $0.125^\circ \times 0.125^\circ$ resolution. We used a median filter over a running box of $21^\circ \times 2^\circ$ to filter out small-scale variability and noise. At this stage we also reduced the resolution to $1^\circ \times 1^\circ$ and 10 d. The data were log-transformed and run through Remez highpass filters¹⁸ in longitude (cut-off length scale 25°) and time (cut-off period 260 d) so as to eliminate seasonal and other large-scale variability.

Gridded sea-level anomalies at $0.25^\circ \times 0.25^\circ$ and 10 d resolution calculated by blending TOPEX/POSEIDON and ERS-2 data were supplied by AVISO (Analysis, Validation and Investigation of Satellite Oceanography). This is an optimally analysed data set, which combines some of the strengths of the two platforms, namely better temporal resolution with TOPEX and smaller cross-track separation with ERS-2. The SSH data were highpass-filtered in the same way as chlorophyll.

We calculated the power spectra of, and the coherence between, chlorophyll and SSH records from time–longitude plots individually for each ocean basin at every 5° of latitude between 40° S and 40° N. We averaged the auto-spectra and the co-spectrum over a 3° latitude band, and smoothed over a moving box of three frequencies and three wavenumbers for statistical convergence. (The equivalent number of degrees of freedom was 9.45 after accounting for the reduction in independence of neighbouring spectral bands due to zero padding and cosine-taper windowing.) Coherence was calculated only where both chlorophyll and SSH power spectral densities were at least 1% of their maximum values.

We used the coherence spectra to construct a filter in the frequency–wavenumber domain. The passband of the filter included only the second and fourth quadrants, where the propagation is westward, and only those wavenumbers and frequencies for which the chlorophyll–SSH coherence was above a threshold of 0.3, which was the 95% confidence limit for significant coherence for the number of degrees of freedom in the spectral analysis¹⁹. This filter was applied to the power spectra of chlorophyll and SSH. The variances of the filtered and unfiltered signals were calculated by integration over wavenumber and frequency.

To estimate the propagation speed of chlorophyll anomalies, we used the ‘Radon’ transform of the power spectrum or of the SSH–chlorophyll coherence. The Radon transform of a two-dimensional image corresponds to the projection of the image intensity along a radial line oriented at a specified angle²⁰. The angle that maximizes the Radon transform is perpendicular to the spectral distribution of the non-dispersive Rossby waves.

The dispersion relation of baroclinic Rossby waves is $\omega = -\beta k(k^2 + \lambda_n^2)^{-1}$, where ω is the natural frequency, β is the rate of change of the Coriolis parameter with meridional displacement and λ_n is the radius of deformation for the n th vertical mode. The radius of deformation is a variable of the Coriolis parameter, density stratification and the total depth of the water column. We used λ_n values calculated in ref. 15 using climatological density profiles.

The phase speed ($C_p = \omega/k$) is nondispersive for long waves and equal to $-\beta\lambda_n^2$. We zonally averaged the radii of deformation, which was originally at $1^\circ \times 1^\circ$ resolution within each basin so as to get one mean phase speed at each latitude.

Received 29 March; accepted 6 November 2000.

- Barber, R. T. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 89–106 (Plenum, New York, 1992).
- Jenkins, W. J. & Goldman, J. C. Seasonal oxygen cycling and primary production in the Sargasso Sea. *J. Mar. Res.* **43**, 465–491 (1985).
- McGillicuddy, D. J. Jr & Robinson, A. R. Eddy-induced nutrient supply and new production in the Sargasso Sea. *Deep-Sea Res.* **44**, 1427–1450 (1997).
- Siegel, D. A., McGillicuddy, D. J. Jr & Fields, E. A. Mesoscale eddies, satellite altimetry and new production in the Sargasso Sea. *J. Geophys. Res.* **104**, 13359–13379 (1999).
- McGillicuddy, D. J. Jr *et al.* Mesoscale variations of biogeochemical properties in the Sargasso Sea. *J. Geophys. Res.* **104**, 13381–13394 (1999).
- McGillicuddy, D. J. Jr, Kosnyrev, V. K., Ryan, J. P. & Yoder, J. A. Covariation of mesoscale ocean color and sea surface temperature patterns in the Sargasso Sea. *Deep-Sea Res.* (in the press).
- Chelton, D. M. & Schlax, M. Global observation of oceanic Rossby waves. *Science* **272**, 234–238 (1996).
- Stammer, D. Global characteristics of ocean variability estimated from regional TOPEX/POSEIDON altimeter measurements. *J. Phys. Oceanogr.* **27**, 1743–1769 (1997).
- Polito, P. & Cornillon, P. Long baroclinic Rossby waves detected by TOPEX/POSEIDON. *J. Geophys. Res.* **102**, 3215–3235 (1997).
- Halliwel, G. R. & Cornillon, P. Westward-propagating SST anomaly features in the Sargasso Sea, 1982–88. *J. Phys. Oceanogr.* **21**, 635–649 (1991).
- Van Woert, M. L. & Price, J. M. Geosat and advanced very high resolution radiometer observations of oceanic planetary waves adjacent to the Hawaiian islands. *J. Geophys. Res.* **98**, 14619–14631 (1993).
- Cipollini, P. *et al.* Concurrent altimeter and infrared observations of Rossby wave propagation near 34° N in the Northeast Atlantic. *Geophys. Res. Lett.* **24**, 889–892 (1997).
- McClain, C. R. *et al.* Science quality SeaWiFS data for global biosphere research. *Sea Technol.* **39**, 10–16 (1998).
- Zang, X. & Wunsch, C. The observed dispersion relationship for North Pacific Rossby wave motions. *J. Phys. Oceanogr.* **29**, 2183–2190 (1999).
- Chelton, D. B. *et al.* Geographical variability of the first baroclinic Rossby radius of deformation. *J. Phys. Oceanogr.* **28**, 433–460 (1998).
- McGillicuddy, D. J. *et al.* Influence of mesoscale eddies on new production in the Sargasso Sea. *Nature* **394**, 263–266 (1998).
- Hayward, T. L. The nutrient distribution and primary production in the central North Pacific. *Deep-Sea Res.* **34**, 1593–1627 (1987).
- IEEE *Programs for Digital Signal Processing*, Algorithm 5.1 (IEEE Press, New York, 1979).
- Emery, W. J. & Thomson, R. E. *Data Analysis Methods in Physical Oceanography* (Pergamon, Oxford/New York, 1997).

20. Lim, J. S. *Two-Dimensional Signal and Image Processing* 42–45 (Prentice Hall, Englewood Cliffs, New Jersey, 1990).

Acknowledgements

We thank the SeaWiFS team for chlorophyll data and AVISO for SSH data; we also thank D. Siegel, P. Polito and S. Schollaert for discussions and recommendations. This work was supported by NASA.

Correspondence and requests for materials should be addressed to B.M.U. (e-mail: m.uz@gso.uri.edu).

Aerogeophysical measurements of collapse-prone hydrothermally altered zones at Mount Rainier volcano

Carol A. Finn[†], Thomas W. Sisson[†] & Maryla Deszcz-Pan^{*}

^{*} US Geological Survey, MS964, PO Box 25046, Denver Federal Center, Denver, Colorado 80225, USA

[†] US Geological Survey, 345 Middlefield Rd, Menlo Park, California 94025, USA

[†] These authors contributed equally to this work.

Hydrothermally altered rocks can weaken volcanoes, increasing the potential for catastrophic sector collapses that can lead to destructive debris flows¹. Evaluating the hazards associated with such alteration is difficult because alteration has been mapped on few active volcanoes^{1–4} and the distribution and severity of subsurface alteration is largely unknown on any active volcano. At Mount Rainier volcano (Washington, USA), collapses of hydrothermally altered edifice flanks have generated numerous extensive debris flows^{5,6} and future collapses could threaten areas that are now densely populated⁷. Preliminary geological mapping and remote-sensing data indicated that exposed alteration is contained in a dyke-controlled belt trending east–west that passes through the volcano’s summit^{3–5,8}. But here we present helicopter-borne electromagnetic and magnetic data, combined with detailed geological mapping, to show that appreciable thicknesses of mostly buried hydrothermally altered rock lie mainly in the upper west flank of Mount Rainier. We identify this as the likely source for future large debris flows. But as negligible amounts of highly altered rock lie in the volcano’s core, this might impede collapse retrogression and so limit the volumes and inundation areas of future debris flows. Our results demonstrate that high-resolution geophysical and geological observations can yield unprecedented views of the three-dimensional distribution of altered rock.

Mount Rainier (Fig. 1), built in the past ~500 kyr atop Tertiary granitic and metavolcanic rocks⁹, is composed of andesite and dacite lavas with subordinate pyroclastic flow deposits¹⁰. Alteration at the summit lies in the volcano’s axial vent system (S, Fig. 1a), through which most magmas that fed the volcano were erupted. Intense alteration (yellow pixels, Fig. 1a) on the volcano’s upper flanks is associated with a zone of open fractures and radial dykes³ (Fig. 1a) emplaced during two episodes of heightened magmatic activity from ~500–400 kyr and ~280–190 kyr¹⁰. Isotopic evidence suggests that most of the alteration was formed where degassing magmas interacted with meteoric water near the surface (R. Rye, personal communication), supporting the geological and geochronological evidence that alteration on the volcano was confined to

episodes of heightened magnetism¹⁰.

During the Holocene, the volcano has generated at least 55 large lahars¹¹, some of which have travelled as far as Puget sound (Fig. 1, inset). Some of the largest collapses, the 5.6 kyr (ref. 6) Osceola that produced a $38 \times 10^8 \text{ m}^3$ mudflow from the east side of the volcano^{6,7}, and the 2.7 kyr Round pass and the Electron⁷ (500 yr ago) that produced $(2\text{--}2.6) \times 10^8 \text{ m}^3$ mudflows from Sunset amphitheater, contain clay and other hydrothermal minerals, and originated by the failure of hydrothermally altered portions of the edifice. The immense Osceola collapse removed the volcano's summit and upper east slope and produced a horseshoe-shaped crater 1.75 km wide (Fig. 1a). Eruptions after 5.6 kyr built a new summit cone within the Osceola crater and covered the upper east slope of the volcano.

Geophysical data have been collected by helicopter at Mount Rainier to discern through-going sections of exposed altered rocks and those obscured beneath snow, vegetation and surficial unaltered rocks. The survey collected four frequencies (837 Hz, 4,341 Hz, 4,737 Hz and 33 kHz) of electromagnetic measurements from a detector towed 45 m above the terrain and total-field magnetic data from a separate detector towed 75 m above the terrain along east- and west-trending lines spaced 250 m apart¹² to obtain information on electrical resistivities and magnetic properties.

Altered volcanic rocks have low electrical resistivities because of their high porosity and clay content, whereas fresh volcanic rocks have high resistivities. Resistivities of unaltered Quaternary volcanic rocks generally exceed $3,000 \Omega \text{ m}$, the Tertiary volcanic rocks range from 1,500 to $3,000 \Omega \text{ m}$, and water-saturated altered rock ranges from ~ 10 to $500 \Omega \text{ m}$ (W. D. Stanley, personal communication).

A least-squares inversion¹³ of the helicopter electromagnetic data¹² produced a three-dimensional model of resistivity as a function of depth. To estimate the thickness and resistivity of the

altered regions, we assigned resistivities of $1 \text{ M}\Omega \text{ m}$ and $3,000 \Omega \text{ m}$ to the ice and unaltered rock, respectively. Owing to a low signal level above the very resistive environment of Mount Rainier, and irregular, high-relief terrain, the inversion only indicated the presence or absence of low resistivity zones and did not provide absolute thickness and resistivity. The maximum depth of penetration of the signal in the least resistive $10 \Omega \text{ m}$ clays is $\sim 50 \text{ m}$ and in unaltered rocks is $\sim 160 \text{ m}$. Altered or fresh rocks lying at greater depths cannot be imaged by the resistivity measurements.

Intense hydrothermal alteration reduces the high magnetization of fresh volcanic rocks¹⁴. Three-dimensional forward modelling shows that uniformly magnetized terrain composed of fresh Quaternary volcanic rocks above their 2,100 m base elevation¹⁵, with a magnetization of 2.5 A m^{-1} in the present Earth's field direction (an inclination of 69° and a declination of 20°) accounts for most of the observed magnetic anomaly. This application of a uniform magnetization in the present Earth's field direction for the Quaternary rocks that compose the topography of Mount Rainier is similar to magnetic models of other unaltered andesitic volcanoes and the magnetization is within a typical range ($1\text{--}4 \text{ A m}^{-1}$)^{14,16}. The Tertiary rocks are nonmagnetic¹⁷. Negative anomalies not correlated with terrain indicate material lacking magnetization. The volume and distribution of these weakly magnetic, presumably altered, masses can be modelled in three dimensions, and their volumes can be approximated if we assume the extreme case of totally nonmagnetic material (0 A m^{-1}).

Intensely altered areas should be both weakly magnetic and of low resistivity. However, modelled nonmagnetic bodies lacking low resistivities can be associated with altered material buried more than 50–160 m beneath ice or relatively unaltered rocks. Nonmagnetic material that coincides with areas of exposed alteration

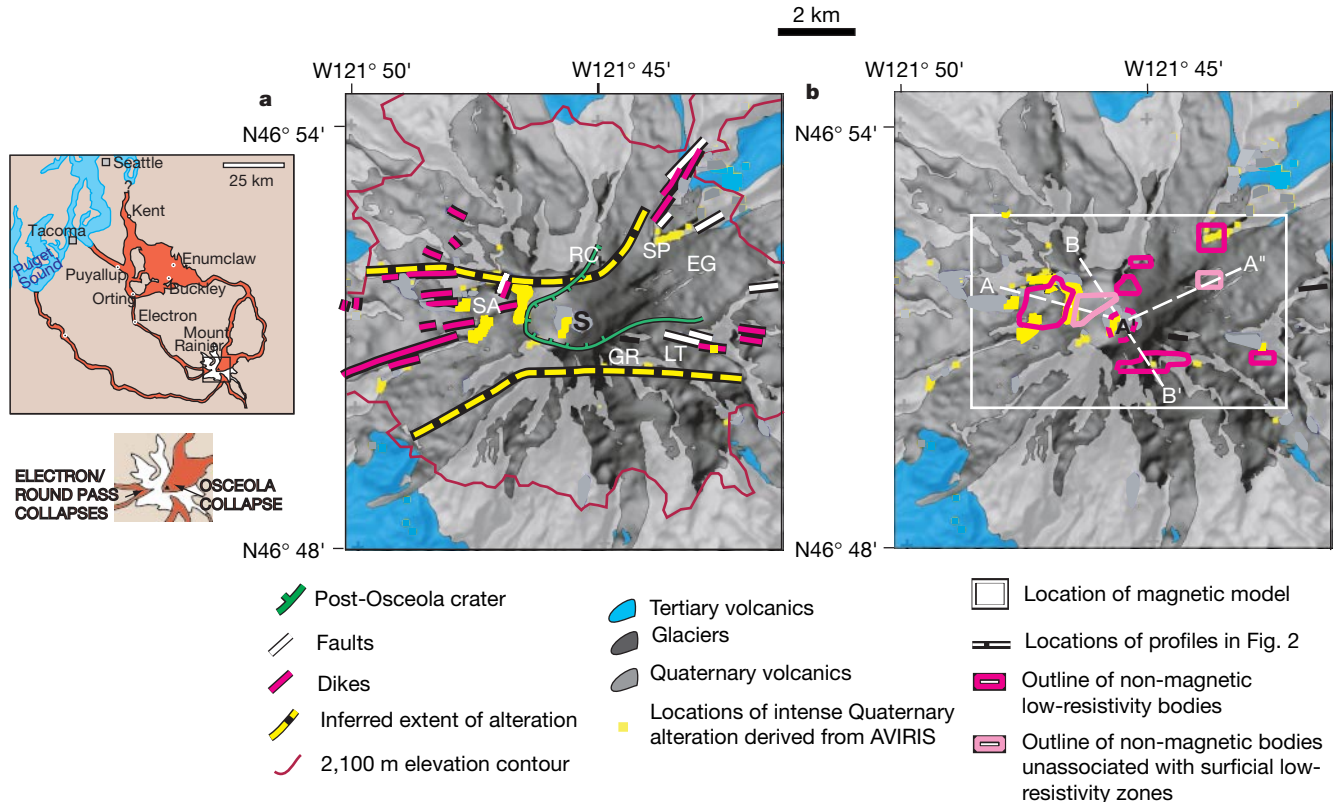


Figure 1 Maps showing the location of mudflows, topography and geology, and geophysical modelling results from Mount Rainier, Washington. Inset, map showing the location of Mount Rainier and Holocene mudflows (in orange). The black box shows the location of **a** and **b**. Topographic map (shading) overlain by generalized geology⁹ and *in situ* alteration mapped on the ground³ and by airborne AVIRIS data⁴. **a**, SA, Sunset

amphitheater; S, summit; RC, Russell cliff; GR, Gibraltar rock; LT, Little Tahoma peak; SP, Steamboat prow; EG, Emmons glacier. **b**, Summary of results of three-dimensional modelling of helicopter-borne electromagnetic and total-field magnetic data¹². The root-mean-square error for the entire three-dimensional magnetic model was 95 nT. AVIRIS, Airborne Visible/Infrared Imaging Spectrometer.

associated with the dyke and fracture system and low resistivities is in Sunset amphitheater, Steamboat prow and south of Little Tahoma peak (Fig. 1). Sunset amphitheater contains the largest thickness of nonmagnetic material (Fig. 2a); the smaller bodies near Steamboat prow and south of Little Tahoma peak are 300–600 m thick. Ice conceals two separate areas of nonmagnetic low-resistivity rocks, 1,500 m thick, below Russell cliff on the northern mouth of the Osceola collapse crater (Fig. 1). On the south margin of the crater, near Gibraltar rock, nonmagnetic low-resistivity altered rocks are buried beneath a few small, altered exposures in largely unaltered rock (Fig. 2b). Between Sunset amphitheater and the summit lies a thick section of nonmagnetic material with low-resistivity zones along the edges only (Figs 1b, 2b); these altered rocks probably shallowly underlie fresh or only lightly altered lavas of the post-Osceola summit cone. Finally, a spatially restricted, but thick, section of nonmagnetic rock without surficial low-resistivity zones lies near small outcroppings of fresh post-Osceola lava in the central Emmons glacier (Figs 1b, 2a) and might reflect altered rocks of the pre-Osceola radial dyke system buried by post-Osceola lavas.

Magnetic modelling was also used to test possible configurations of altered rock at the summit, in the central core of the volcano and under the eastern side of the volcano. These areas have significance for assessments of edifice stability and volcano hazards because extensive edifice core alteration would represent enormous volumes of potentially weak rock, and extensive, shallow east-slope alteration could promote large, east-directed edifice failures. Many workers have postulated that a core of highly altered rock underlies the summit and extends through the edifice, on the basis of highly altered rocks exposed at the summit crater^{2,3,8} and high proportions of highly altered rocks mobilized by the major collapse of the Osceola edifice^{5,7}. A low-amplitude (100–200 nT) magnetic low over the summit (Fig. 2) does in fact correspond to a low-resistivity zone and mapped intense alteration (Fig. 1). However, modelling of the magnetic data shows that exposed intense alteration is surficial and cannot be more than 20–50 m thick (Fig. 2). A large (>500 m in diameter) buried central core of highly altered, nonmagnetic rock is also excluded by the modelling, suggesting that any coherent masses

of intensely altered rock must be very small and lie far beneath the summit (>1,000 m), if it is present at all. Intensely altered fracture coatings or small pockets cannot be ruled out by the geophysical data and might account for the compositions of summit fumarole gases⁸. Intensely altered material, thinly mantled by post-Osceola lava flows and ice, was thought to compose the upper east side of the volcano^{2,3} and to represent a potential failure surface for large eastwards-directed collapses. However, the magnetic signal of the upper east slope can be modelled as being due entirely to unaltered lavas. These results indicate that most of the summit, core and upper east side of the volcano are free of weak, highly altered rock and have a relatively low risk of failure¹⁸.

The modelled uneven distribution of thick, intense alteration differs from the simple east–west band extrapolated from surficial exposures^{2,3} (yellow line, Fig. 1a) and the clay content of debris flows. The greatest volume of nonmagnetic low-resistivity material, known from exposures to be intensely altered rock, lies on the upper west side of the volcano and constitutes roughly $16 \times 10^8 \text{ m}^3$, of which $9 \times 10^8 \text{ m}^3$ underlie Sunset amphitheater and $6 \times 10^8 \text{ m}^3$ lie immediately to the east, towards the summit. Smaller volumes of $\sim 3 \times 10^8 \text{ m}^3$ lie below Russell cliff, $6 \times 10^8 \text{ m}^3$ under the Gibraltar rock area and $0.07 \times 10^8 \text{ m}^3$ under the summit.

These general volume estimates based on the geophysical data show that Mount Rainier's upper west side contains the largest remaining mass of highly altered rocks on the volcano. This means that the Electron and Round pass mudflows, which originated from that area (Fig. 1, inset), left a substantial quantity of altered material at great height, setting the stage for future debris flows. The volume of $16 \times 10^8 \text{ m}^3$ for hazardous rock on the west side is larger than all except the Osceola mudflow, indicating that sufficient altered material remains in the Sunset amphitheater area to yield future debris flows at least the size of the Electron event and capable of reaching areas that are now densely populated.

Before the Osceola collapse event, the central core and upper east slope of the volcano were highly altered as supported by the abundant alteration minerals within the Osceola deposit, by the continuity of dykes, fractures and associated alteration on the west

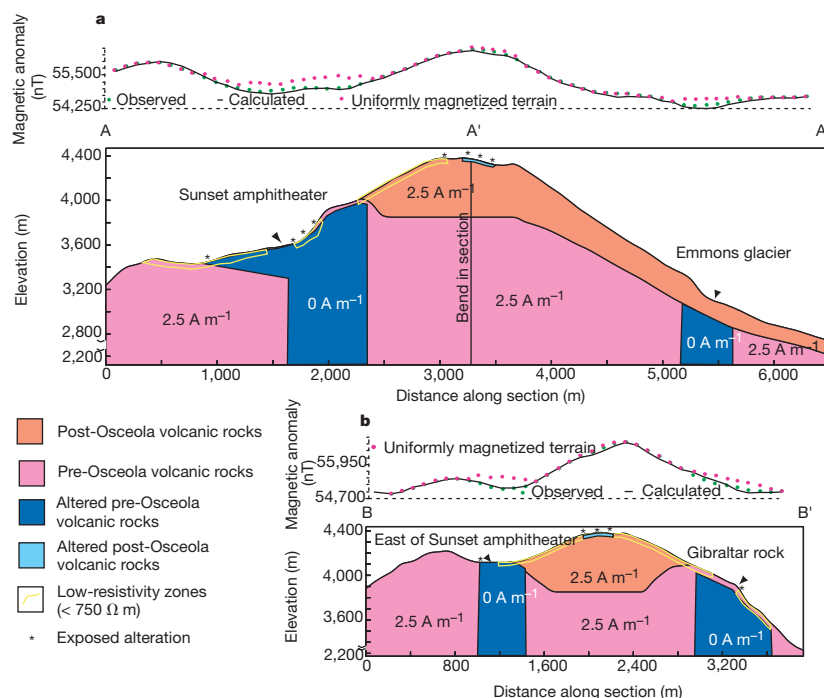


Figure 2 Cross-sections derived from geology¹⁸ and the three-dimensional magnetic and electrical resistivity models. **a**, **b**, Profile locations are shown in Fig. 1b. The magnetic anomalies due to uniformly magnetized terrain were calculated from the same terrain

model and magnetization as the unaltered portions of the non-uniform model (an inclination of 69°, a declination of 20° and an intensity of 2.5 A m^{-1}).

and lower east sides of the volcano, and by the rapid transformation of the collapsed material into a far-travelled lahar. The three large, thick, nonmagnetic bodies that ring the edge of the Osceola palaeocrater (below Russell cliff, east of Sunset amphitheater and near Gibraltar rock) (Fig. 2) are probably remnants of the old altered core of the volcano. The absence of thick altered zones beneath the modern summit and the upper east slope suggests not only that the Osceola collapse removed the altered core and upper eastern portion of the old dyke system (Fig. 1a) to substantial depths, but also that the vertical depth of incision of the Osceola failure might have been limited by the base of highly altered rock.

The absence of a large volume of alteration beneath the modern summit and east slope might restrict the collapse of altered material to the west side of the volcano¹⁸, suggesting that a mudflow event as large as the Osceola is no longer likely. If collapse retrogresses into the core of the volcano, the relatively coherent core material might generate a debris avalanche that would be far less mobile than clay-rich lahars. As alteration is associated primarily with eruptive periods at Mount Rainier¹⁰, the development of future weak, altered zones might depend on the frequency and volume of eruptions. The 25–50 m thickness of alteration at the modern summit has formed since ~2,000–5,000 yr ago. If magmatism and alteration were to continue at these Holocene rates, it would take at least 20,000 yr to alter an appreciable thickness (>500 m) of the volcano's core. Significant alteration associated with dyke injection also takes 50–100 kyr¹⁰.

This first detailed assessment of the internal distribution of altered zones in an active volcano, using geophysical measurements, differs substantially from the distribution extrapolated from surficial exposures alone^{2–4}. Lahars generated by the collapse of structurally incompetent hydrothermally altered rock are most probable on the west side of the volcano¹⁸. Strong shaking of the edifice during even small eruptive events could dislodge altered rock and generate a lahar capable of reaching densely populated areas. Although edifice collapse does not require weakened altered rocks, the widespread preservation of old (100–200 kyr) lava flows at high elevations on Mount Rainier¹⁰, as well as the scarcity of debris avalanche deposits, as opposed to lahar deposits, suggests that unaltered flanks collapse infrequently. But Mount Rainier has produced numerous far-travelled lahars that contain little or no altered material. Some of these alteration-free lahars probably formed as pyroclastic flows or disaggregating active lava flows that swept across and incorporated glacial ice. Lahars originating by this magma–ice interaction threaten all valleys draining the edifice. Nevertheless, the collapse of altered flanks, either during or independently of eruptive activity, is a primary hazard at Mount Rainier and elsewhere, and high-resolution geophysical surveys interpreted with the benefit of detailed geological mapping is an effective tool for evaluating, substantiating and quantifying hazards from collapse-generated debris flows. □

Received 5 June; accepted 13 November 2000.

1. Lopez, D. L. & Williams, S. N. Catastrophic volcanic collapse; relation to hydrothermal processes. *Science* **260**, 1794–1796 (1993).
2. Frank, D. Surficial extent and conceptual model of hydrothermal system at Mount Rainier, Washington. *J. Volcanol. Geotherm. Res.* **65**, 51–80 (1995).
3. Zimbelman, D. R. *Hydrothermal Alteration and its Influence on Volcanic Hazards; Mount Rainier, Washington, a Case History* (Univ. Colorado, Boulder, Colorado, 1996).
4. Crowley, J. K. & Zimbelman, D. R. Mapping hydrothermally altered rocks on Mount Rainier, Washington, with Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data. *Geology* **25**, 559–562 (1997).
5. Crandell, D. R. & Waldron, H. H. A Recent volcanic mudflow of exceptional dimensions from Mount Rainier, Washington. *Am. J. Sci.* **254**, 349–362 (1956).
6. Vallance, J. W. & Scott, K. M. The Osceola Mudflow from Mount Rainier: sedimentology and hazard implications of a huge clay-rich debris flow. *Geol. Soc. Am. Bull.* **109**, 143–163 (1997).
7. Scott, K. M., Vallance, J. W. & Pringle, P. T. Sedimentology, behavior, and hazards of debris flows at Mount Rainier, Washington. *US Geol. Surv. Prof. Pap.* **1547**, 1–56 (1995).
8. Moran, S. C., Zimbelman, D. R. & Malone, S. D. A model for the magmatic hydrothermal system at Mount Rainier, Washington, from seismic and geochemical observations. *Bull. Volcanol.* **61**, 425–436 (2000).

9. Fiske, R. S., Hopson, C. A. & Waters, A. C. Geology of Mount Rainier National Park Washington. *US Geol. Surv. Prof. Pap.* **444**, 93 (1963).
10. Sisson, T. W. & Lanphere, M. A. Geologic controls on the timing and location of flank alteration at Mt. Rainier, Washington. *Eos* **80**, F1151–F1152 (1999).
11. Crandell, D. R. Postglacial lahars from Mount Rainier Volcano, Washington. *US Geol. Surv. Prof. Pap.* **677**, 75 (1971).
12. Rystrom, V. L., Finn, Carol A. & Descz-Pan, Maryla High resolution, low altitude aeromagnetic and electromagnetic survey of Mt Rainier. *US Geological Survey Open-File Report 00-027* [online], (<http://greenwood.cr.usgs.gov/pub/open-file-reports/ofr-00-0027/Rainierwebpage.html>) (2000).
13. Descz-Pan, M., Fitterman, D. V. & Labson, V. F. Reduction of inversion errors in helicopter EM data using auxiliary information. *Explor. Geophys.* **29**, 142–146 (1998).
14. Woodward, D. J. & Mumme, T. C. Variation of magnetisation on White Island, New Zealand. *N.Z. J. Geol. Geophys.* **36**, 447–451 (1993).
15. Fiske, R. S., Hopson, C. A. & Waters, A. C. *Geologic Map and Section of Mount Rainier National Park Washington* (US Geological Survey Miscellaneous Investigations Series I-432, 1964).
16. Finn, C. & Williams, D. L. An aeromagnetic study of Mount St. Helens. *J. Geophys. Res.* **92**, 10194–110206 (1987).
17. Williams, D. L. & Finn, C. A. Evidence for a shallow pluton beneath the Goat Rocks Wilderness, Washington, from gravity and magnetic data. *J. Geophys. Res.* **92**, 4867–4880 (1987).
18. Reid, M. E., Christian, S. B., Brien, D. L. & Sisson, T. W. 3-D gravitational stability of stratovolcanoes. *Eos* **80**, F1151 (1999).

Acknowledgements

We thank D. Fitterman, V. J. S. Grauch and P. Lipman for helpful reviews. This work was supported by the Mineral Resource and Volcano Hazards Programs of the US Geological Survey.

Correspondence and requests for materials should be addressed to C.A.F. (e-mail: cfinn@usgs.gov).

Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton

Purificación López-García*, Francisco Rodríguez-Valera*, Carlos Pedrós-Alíó† & David Moreira*

* División de Microbiología, Universidad Miguel Hernández, 03550 San Juan de Alicante, Spain

† Institut de Ciències del Mar, CSIC, 08039 Barcelona, Spain

Phylogenetic information from ribosomal RNA genes directly amplified from the environment changed our view of the biosphere, revealing an extraordinary diversity of previously undetected prokaryotic lineages. Using ribosomal RNA genes from marine picoplankton, several new groups of bacteria and archaea have been identified, some of which are abundant^{2–4}. Little is known, however, about the diversity of the smallest planktonic eukaryotes, and available information in general concerns the phytoplankton of the euphotic region. Here we recover eukaryotes in the size fraction 0.2–5 µm from the aphotic zone (250–3,000 m deep) in the Antarctic polar front. The most diverse and relatively abundant were two new groups of alveolate sequences, related to dinoflagellates that are found at all studied depths. These may be important components of the microbial community in the deep ocean. Their phylogenetic position suggests a radiation early in the evolution of alveolates.

We amplified 18S rRNA genes from samples taken at 250, 500, 2,000 and 3,000 m deep at the Antarctic polar front limit in a transect along the Drake passage (59° 19' 48" S, 55° 45' 11" W, sea floor at 3,671 m). This sampling site interested us because it is a region of water-mass mixing from the Atlantic and Southern oceans. It corresponds to cold and oligotrophic waters where microbial biomass, especially at 3,000 m deep, reached minimal values in the area as deduced from DNA yields (see Methods). We constructed 18S rRNA environmental gene libraries from the 0.2–

5 μm planktonic fraction and, for comparison, also from the microbial fraction $>5 \mu\text{m}$ at 3,000 m deep. After partial sequencing of the 3' region of the gene (700 base pairs, bp, on average), BLAST searches⁵ and phylogenetic reconstruction distance methods provided us with a first survey of the type of eukaryotic sequences present in our samples. Twenty-four representative clones from all depths were subsequently chosen for complete sequencing. The complete sequences were aligned with 1,443 additional 18S rRNA gene sequences retrieved from databanks. A subset of 101 complete sequences was then selected for phylogenetic analysis, taking special care to include a taxonomically broad sample of eukaryotes (all and closest relatives to our sequences) to minimize artefacts related to taxonomic sampling. We constructed distance (neighbour-joining, NJ), maximum-parsimony (MP) and maximum-likelihood (ML) trees, which produced similar congruent results. Figure 1 shows an ML tree displaying the eukaryotic microbial diversity found.

As was expected of cold, highly oxygenated waters, the majority of sequences affiliate with the eukaryotic 'crown', the densely branched apical part of the eukaryotic tree⁶. However, we found three phylotypes belonging to the early branching part of the 18S rRNA tree (Fig. 1a). DH148-5-EKD18 represents a new lineage emerging in the region of the Archezoa. The large length of its branch suggests that it could correspond to a parasite whose rRNA has evolved rapidly. It would thus be 'attracted' to the base of the tree by a long-branch attraction artefact, as indeed occurs with Microsporidia⁷. This may be supported by the occurrence of several specific deletions in this sequence (data not shown), as fast-evolving eukaryotic sequences are often characterized by length variation⁸. DH145-EKD11 also corresponds to a new eukaryotic lineage of uncertain phylogenetic ascription, although it emerges in a region of the tree occupied by amoeboid organisms (*Phreatamoeba*, *Entamoeba* or Myxozoa) (Fig. 1a). We could retrieve very few sequences with the primer set EK-1F + EK-1520R, which was used only with the 3,000-m sample, but all of them (here represented by DH148-EKB1) were related to *Diplonema* spp., which are euglenozoan heterotrophs frequently found in marine benthic sites⁹.

The diversity of crown eukaryotes is much larger (Fig. 1b). The most frequently retrieved groups were the alveolates, followed by heterokonts. We also found sequences belonging to fungi, and to the amoeboid phagotrophic acantharean radiolaria. Fungi reaffirm themselves as one of the most ecologically successful eukaryotic lineages; they have even been isolated from the bottom of the Mariana trench (10,897 m)¹⁰. Within the heterokonts, we detected sequences related to the labyrinthulids (DH147-EKD10). These are relatively common in the sea and play a role in decomposition processes colonizing faecal pellets also under deep-sea conditions¹¹. Two other sequences, DH148-5-EKD53 and DH144-EKD10, do not clearly affiliate with any known species, and may represent new lineages of heterokonts. We also retrieved a pennate diatom sequence (DH148-5-EKD54) at 3,000 m that could correspond to a sinking cell. However, as it is very similar to *Pseudo-nitzschia* spp. sequences, common dinoflagellate endosymbionts, this sequence could instead derive from a dinoflagellate endosymbiont.

Alveolate sequences were by far the most diverse in our samples. Within the commonly predatory ciliates, we obtained new oligohymenophorean (DH148-5-EKD6) and colpodean (DH147-EKD23) sequences. We also recovered typical dinoflagellate sequences from all depths and both planktonic fractions (Figs 1b and 2). These sequences are related to Gymnodiniales (often lacking a theca cell wall) and Prorocentrales (theca with two plates), which are usually small¹². Interestingly, the vast majority of sequences obtained grouped in two major clades between dinoflagellates and apicomplexans (we have termed these marine alveolate groups I and II). These sequences were mainly retrieved from the smallest planktonic fraction at all depths (Figs 1b and 2). In terms of genetic divergence, the diversity found within these groups is equivalent to

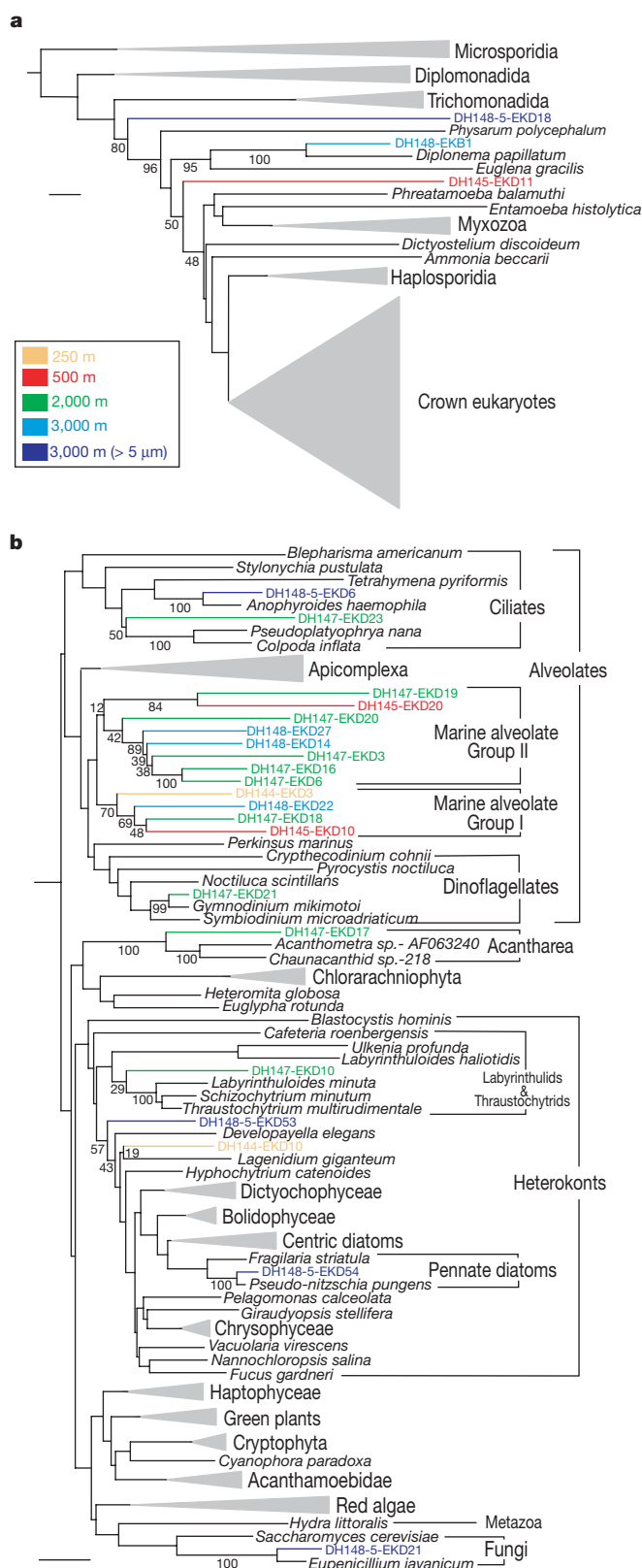


Figure 1 Maximum-likelihood tree of eukaryotic phylotypes in deep Antarctic waters constructed using 101 eukaryotic 18S rRNA sequences. The tree has been split in two parts representing the basal part (a) and the crown (b) of the eukaryotic rRNA-based phylogeny. The outgroup branch (archaea) is not shown. Thin triangles correspond to two representative species of a given taxon, three in the case of Apicomplexa. Bootstrap values are given only below nodes concerning the new eukaryotic sequences. The colour code indicates sea depths at which sequences were obtained. Scale bars correspond to 15 substitutions per 100 positions for a unit branch length.

that displayed by all dinoflagellates known to date. Taking into account that these sequences come from the aphotic region of a single (low biomass) sampling site, this diversity is astonishing. On the other hand, many microbial groups seem to be ubiquitous in the ocean, probably as a result of current mixing¹³, and this could also be true of the newly discovered alveolate lineages. In fact, Guillou *et al.*¹⁴ retrieved a picoplanktonic sequence (OLI2001) from 100 m deep in the equatorial Pacific that emerged at the base of the (only three) dinoflagellate sequences they used. When we included it in our larger alignment, OLI2001 branches within the marine alveolate group I (Fig. 2). This confirms the observation of this group at upper parts of the water column, and suggests the ubiquity of this lineage in the sea. The discovery of these two new clusters of sequences branching amid apicomplexans and dinoflagellates sup-

ports the idea of an early radiation in the evolutionary history of alveolates. Indeed, internal bootstrap values within alveolates are quite small in our analysis, with the exception of ciliates (76%) and marine alveolates group I (70%) (Fig. 1b). This trend may be diagnostic for radiation processes where the order of branch emergence is difficult to assess¹⁵.

We do not observe a clear pattern of diversity fractionation with depth in the range studied (250–3,000 m), especially for marine alveolates group I and II (Fig. 2). This is consistent with the homogeneity of the physico-chemical conditions in this part of the aphotic water column (poor nutrient concentration, absence of light, average temperature around 2°C). These novel groups accounted for most of the diversity found, and corresponded to 65 to 76% of the sequences retrieved from the smallest planktonic

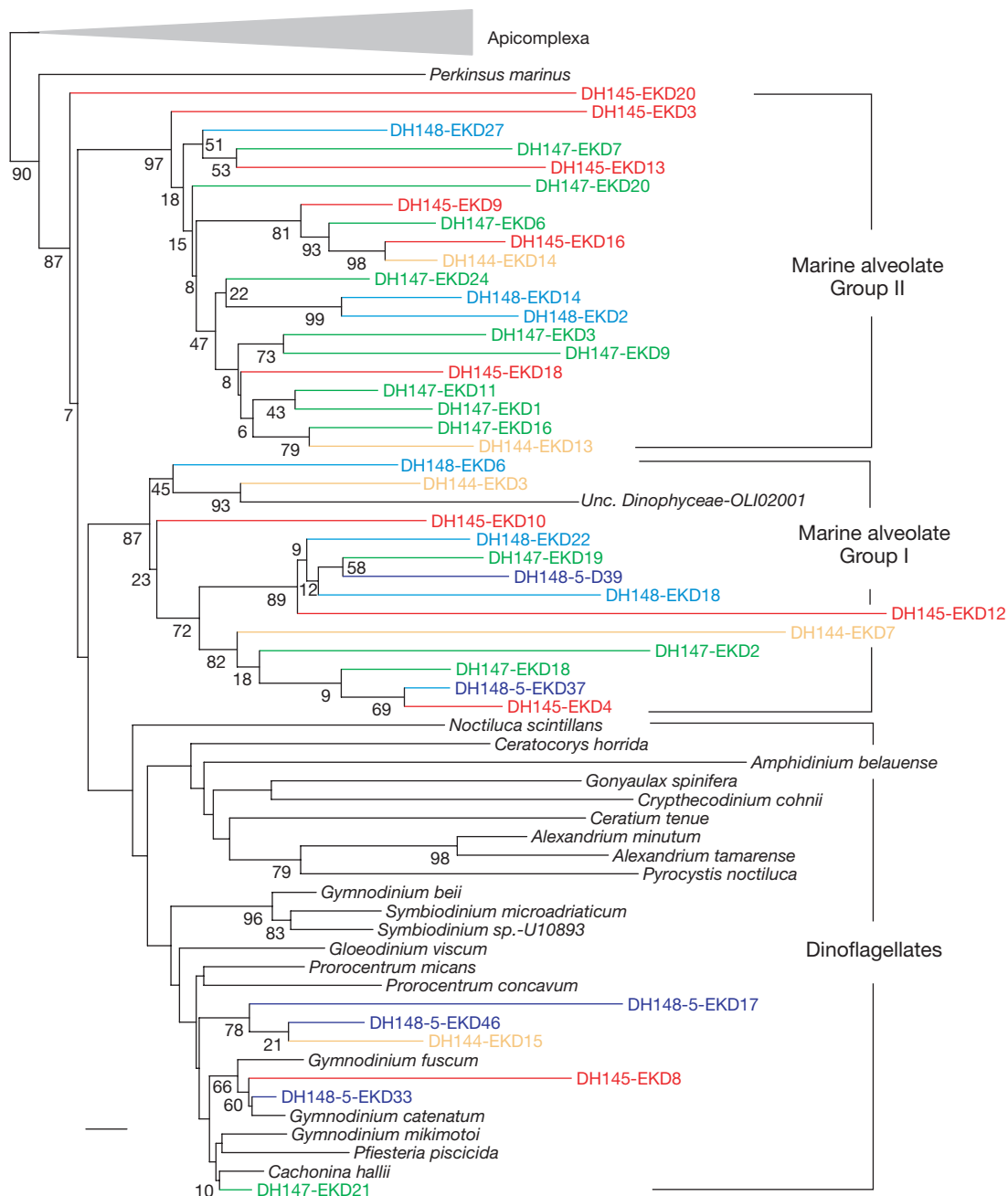


Figure 2 Maximum-likelihood tree showing the diversity of new dinoflagellates and marine alveolate groups I and II. The tree was constructed using partial 3' end 18S rRNA sequences. Bootstrap values corresponding to the new eukaryotic sequences and also

those above 80% are given below the respective nodes. The scale bar corresponds to 5 substitutions per 100 positions for a unit branch length. The colour code is as in Fig. 1.

fraction at different depths, suggesting that they are also relatively abundant. By contrast, we could detect a significant difference between the eukaryotic diversity observed in the fractions 0.2–5 μm and >5 μm at 3,000 m deep. Sequences recovered from the larger fraction were more diverse, and marine alveolate groups I and II accounted only for 31% of the clones. This trend was compensated by a larger proportion of heterokonts, and other groups such as fungi or even animals—a relatively large number of copepod sequences was amplified (data not shown). We did not identify foraminifera in our samples, although some of the basal diverging lineages could belong to these organisms. Sequences from planktonic foraminifera are scarce in databanks, and they are very divergent, often containing large insertions¹⁶.

The diversity of small eukaryotes found at high ocean depth (250–3,000 m) in the Antarctic appears to be at least as important as that of prokaryotes, including both archaea and bacteria (unpublished results). They may reach the deeper parts of the water column by sinking from the upper region, where primary production, diversity and nutrient concentration are higher¹⁷. However, we did not find sequences corresponding to several phototrophic groups that are common in surface waters (such as haptophytes and diverse diatoms); this suggests that the new lineages may be thriving in the deep ocean. A small planktonic size would also be in agreement with life in cold and oligotrophic waters¹⁷.

Our findings may be of interest for at least three disciplines. First, for ecology and biogeochemistry, as until now most measurements of prokaryotic carbon cycling in the sea have been based on size-exclusion techniques that cannot discriminate between prokaryotes and very small eukaryotes¹⁷. Second, for phylogeny and evolution. The two new alveolate groups support an early radiation within this phylum and could represent intermediate taxa between dinoflagellates and apicomplexans. These could be used to test some of the evolutionary scenarios proposed (see ref. 18 for example). For general eukaryotic phylogeny, the new lineages in the basal region of the tree may help to break long branches, stabilizing this problematic region¹⁵. Furthermore, our present knowledge of eukaryotic diversity is based on phenotypical traits and species isolation but, for very small eukaryotes, tiny sizes may have precluded their recognition as such by direct observation¹. Molecular ecology techniques based on rRNA amplification could thus stimulate changes in protistology like those that occurred in prokaryotic microbiology. Finally, the existence of a wide variety of small eukaryotes should be considered for general micropalaeontological studies. Easily fossilizable forms (like some dinoflagellates), may have led to the misinterpretation of some microfossils as prokaryotes. This could partly explain problems such as the gap existing between the earliest unequivocal occurrence of dinoflagellates (~240 million years ago) and that of dinoflagellate biomarkers, dinosteranes (~520 million years ago)¹⁹, as well as the decline in the dinoflagellate fossil record observed during the Tertiary period²⁰. □

Methods

Experimental methods

Samples were collected in Niskin bottles on December 1998, during the Spanish oceanographic campaign DHARMA 98 (He052; <http://www.ugbo.csic.es>). For the samples used in this work, volumes from 20 to 341 of sea water from 250, 500, 2,000 and 3,000 m deep were prefiltered through a nylon mesh, filtered through a 5- μm pore-size filter, and the remaining plankton collected in 0.2- μm Sterivex filters (0.2–5 μm fraction). After a proteinase K–SDS lysis step, nucleic acids were extracted as previously described²¹ with a yield of 0.327, 0.146, 0.169, and 0.068 μg DNA per litre of sea water, respectively. 18S rRNA genes presented in this work were amplified by polymerase chain reaction (PCR) using the specific primers EK-1F (CTGGTTGATCCTGCCAG), EK-82F (GAAACTGCG AATGGCTC) and EK-1520R (CYGCAGGTTCACTAC) under previously described conditions⁴. Four additional specific eukaryotic primer sets tested gave similar results to the combination EK-82F/EK-1520R. rDNA clone libraries were constructed using the Topo TA Cloning system (Invitrogen). After plating, 24 to 104 positive transformants per library were screened by PCR amplification of inserts using flanking vector primers. Expected-size amplicons were subsequently cleaned using the QIAquick PCR purification system (Qiagen). Cleaned PCR products were directly partially sequenced in an ABI Prism

377 apparatus (Perkin Elmer Applied Biosystems) using the ABI Prism dRhodamine terminator cycle sequencing ready reaction kit with either primer EK-1F or EK-82F. After preliminary phylogenetic analysis, 13 dinoflagellate-related marine alveolate clones, and 11 eukaryotic clones of different phylogenetic affiliation were chosen for complete sequencing. Inserts were sequenced twice using both flanking vector primers and specific eukaryotic primers. Specific internal primers DIN-1F (GTTGTTGCGGTTAAAAAGC), for dinoflagellate-related clones, and EK-555F (AGTCTGGTGCCAGCAGCCG) and EK-1269R (AAGAACGGCCATGCACCAC), for eukaryotes, were designed to complete and overlap central insert sequences.

Phylogenetic analyses

1,443 eukaryotic 18S rRNA sequences were retrieved from GenBank and the rRNA Database at the University of Antwerp (<http://rrna.uia.ac.be/>). They were aligned together with the Antarctic clone sequences using CLUSTAL W²², and the resulting multiple alignment was manually edited using the program ED from the MUST package²³. Partial NJ trees were constructed for the different eukaryotic taxa to choose a representative subset of 101 sequences, avoiding partial and fast-evolving ones, for further phylogenetic analyses (Fig. 1). Fifteen archaeal sequences were included as the outgroup. Gaps and ambiguously aligned positions were excluded from our analyses, resulting in an alignment of 1,275 positions. An additional alignment including 39 partial, new Antarctic alveolate sequences (580 unambiguous positions) was constructed to analyse the phylogeny of the two new alveolate groups (Fig. 2). MP and ML trees were constructed, respectively, with the programs PAUP 3.1 (ref. 24) and NUCML from the MOLPHY 2.3 package²⁵ using a heuristic quick-add OTUs search and default values. Bootstrap proportions were estimated using 1,000 replicates for NJ and MP trees, and using the REL method²⁶ on the 1,000 top-ranking trees for ML trees. The alpha parameter value of the gamma distribution accounting for among-site rate variation was computed using PUZZLE²⁷. Alignments, trees and list of species used are available on request.

Received 1 August; accepted 6 November 2000.

- Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740 (1997).
- Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **356**, 148–149 (1992).
- Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaeobacterial group from marine plankton. *Nature* **356**, 148–149 (1992).
- DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
- Altschul, S. F. & Koonin, E. V. Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem. Sci.* **23**, 444–447 (1998).
- Knoll, A. H. The early evolution of eukaryotes: a geological perspective. *Science* **256**, 622–627 (1992).
- Embley, T. M. & Hirt, R. P. Early branching eukaryotes? *Curr. Opin. Genet. Dev.* **8**, 624–629 (1998).
- Stiller, J. & Hall, B. Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol. Biol. Evol.* **16**, 1270–1279 (1999).
- Larsen, J. & Patterson, J. L. Some flagellates (Protista) from tropical marine sediments. *J. Nat. Hist.* **24**, 801–937 (1990).
- Takami, H., Inoue, A., Fuji, F. & Horikoshi, K. Microbial flora in the deepest sea mud of the Mariana Trench. *FEMS Microbiol. Lett.* **152**, 279–285 (1997).
- Raghukumar, S. & Raghukumar, C. Thraustochytrid fungoid protists in faecal pellets of the tunicate *Pegea confederata*, their tolerance to deep-sea conditions and implication in degradation processes. *Mar. Ecol. Prog. Ser.* **190**, 133–140 (1999).
- Lenaers, G., Scholten, C., Bhaud, Y., Saint-Hilaire, D. & Herzog, M. A molecular phylogeny of dinoflagellate protists (pyrrhophyta) inferred from the sequence of 24S rRNA divergent domains D1 and D8. *J. Mol. Evol.* **32**, 53–63 (1991).
- Darling, K. F. *et al.* Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature* **405**, 43–47 (2000).
- Guillou, L., Moon-Van Der Staay, S. Y., Claustre, H., Partensky, F. & Vaulot, D. Diversity and abundance of Bolidophyceae (Heterokonta) in two oceanic regions. *Appl. Environ. Microbiol.* **65**, 4528–4536 (1999).
- Philippe, H. *et al.* Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc. R. Soc. Lond. B* **267**, 1213–1221 (2000).
- Pawlowski, J. *et al.* Extreme differences in rates of molecular evolution of foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. *Mol. Biol. Evol.* **14**, 498–505 (1997).
- Fenchel, T., King, G. M. & Blackburn, T. H. *Bacterial Biogeochemistry: The Ecophysiology of Mineral Cycling* (Academic, London, 1998).
- Cavalier-Smith, T. Kingdom protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–994 (1993).
- Moldovan, J. M. & Talyzina, N. M. Biogeochemical evidence for dinoflagellate ancestors in the early cambrian. *Science* **281**, 1168–1170 (1998).
- McRae, R. A., Fensome, R. A. & Williams, G. L. Fossil dinoflagellate diversity, origins and extinctions and their evolutionary significance. *Can. J. Bot.* **74**, 1987–1994 (1996).
- Massana, R., Murray, A. E., Preston, C. M. & DeLong, E. F. Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl. Environ. Microbiol.* **63**, 50–56 (1997).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Philippe, H. MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res.* **21**, 5264–5272 (1993).
- Swofford, D. L. PAUP: phylogenetic analysis using parsimony, version 3.1.1 (Illinois Natural History Survey, Champaign, 1993).
- Adachi, J. & Hasegawa, M. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* **28**, 1–150 (1996).
- Kishino, H., Miyata, T. & Hasegawa, M. Maximum likelihood inference of protein phylogeny, and the origin of chloroplasts. *J. Mol. Evol.* **31**, 151–160 (1990).

27. Strimmer, K. & von Heseler, A. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**, 964–969 (1996).

Acknowledgements

We thank A. Roger for critical reading of the manuscript, A. López-López for DNA extraction and M. L. Campos for sequencing help. This work was supported by the European MIDAS project. The Hesperides campaign DHARMA98 was financed by the Spanish Research Council (CSIC). P.L.G. was financed by a postdoctoral contract of the Spanish Ministerio de Educación y Cultura. Sequences have been deposited in GenBank under accession numbers AF290036 to AF290085.

Correspondence and requests for materials should be addressed to D.M.
(e-mail: david.moreira@umh.es).

Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity

Seung Yeo Moon-van der Staay[†], Rupert De Wachter[‡] & Daniel Vaulot^{*‡}

^{*} Station Biologique, UPR 9042 Centre National de la Recherche Scientifique et Université Pierre et Marie Curie, BP 74, 29682 Roscoff Cedex, France

[‡] Departement Biochemie, Universiteit Antwerpen (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium

Picoplankton—cells with a diameter of less than 3 μm —are the dominant contributors to both primary production and biomass in open oceanic regions^{1,2}. However, compared with the prokaryotes³, the eukaryotic component of picoplankton is still poorly known. Recent discoveries of new eukaryotic algal taxa based on picoplankton cultures^{4,5} suggest the existence of many undiscovered taxa. Conventional approaches based on phenotypic criteria have limitations in depicting picoplankton composition due to their tiny size and lack of distinctive taxonomic characters⁶. Here we analyse, using an approach that has been very successful for prokaryotes⁷ but has so far seldom been applied to eukaryotes⁸, 35 full sequences of the small-subunit (18S) ribosomal RNA gene derived from a picoplanktonic assemblage collected at a depth of 75 m in the equatorial Pacific Ocean, and show that there is a high diversity of picoeukaryotes. Most of the sequences were previously unknown but could still be assigned to important marine phyla including prasinophytes, haptophytes, dinoflagellates, stramenopiles, choanoflagellates and acantharians. We also found a novel lineage, closely related to dinoflagellates and not previously described.

A sequence search with the EMBL gene databank showed that only 2 of the 35 18S rDNA sequences (OLI11030 and OLI11015) from the Pacific Ocean had significant identity (more than 99%) to known sequences: the ubiquitous picoplanktonic species *Pelagomonas calceolata* and a recently sequenced acantharian⁹, respectively. The maximum sequence identities of the other environmental sequences to known eukaryote 18S rDNAs ranged from 82% to 97%. The global phylogenetic tree (Fig. 1) obtained with both environmental clones and available sequences is largely congruent with those found previously^{10–12}, although many evolutionary relationships between the eukaryotic crown taxa are not clear, and the bootstrap values at the nodes are low, as indicated in earlier studies. The phylogenetic positions of the environmental clones are also supported by detailed, separate phylogenetic analyses of sub-

groups, using more 18S rDNA sequences available (data not shown). The presence of sequences from lineages that are known to harbour picoplanktonic representatives, such as the prasinophytes or the pelagophytes (Fig. 1), and the converse absence of sequences from larger cells such as diatoms despite their ubiquity in Pacific waters¹³, confirms that the approach taken specifically targeted picoplankton.

Most of the work on oceanic eukaryotic picoplankton has focused on its photosynthetic component because chlorophyll fluorescence makes it easy to detect by flow cytometry¹ and pigment signatures permit inferences to be made about its taxonomic composition. Among autotrophs, haptophytes constitute one of the major picoplanktonic lineages, as suggested by the dominance of the diagnostic carotenoid 19'-hexanoyloxyfucoxanthin in most oceanic waters¹⁴. Indeed, four haptophyte clones (OLI11056, OLI11019, OLI11072 and OLI11007) were observed in the Pacific sample. Separate phylogenetic analyses of these environmental clones, adding more haptophyte sequences, suggest that clones OLI11072 and OLI11019 are more specifically related to *Chrysochromulina leadbeateri*. Clone OLI11007 belongs to a recently revealed environmental lineage that is related either to coccolithophorids or to *Phaeocystis*¹⁴. Clone OLI11056 forms a somewhat independent clade as a sister to the *Chrysochromulina* clade and that uniting the *Prymnesium*, *Imantonia* and part of the *Chrysochromulina* species (data not shown).

Another key group in the picoplanktonic autotrophs is the prasinophytes, primitive green algae that have been repeatedly isolated from marine waters⁶. Indeed, three clones (OLI11059, OLI11305 and OLI11345) were assigned to this class (Fig. 1). Phylogenetic analyses strongly support the affinity between OLI11059 and the unidentified coccoid prasinophyte CCMP 1205, whereas the other two sequences seem to form a new clade, not yet represented in culture. The exact branching order of the lineage leading to these clones and CCMP 1205 in other early diverging prasinophyte lineages is still not clear (bootstrap value <50%).

Stramenopiles or heterokonta contain key oceanic algal classes, in particular the ubiquitous diatoms, but also heterotrophic groups such as the bicoseoids. Clone OLI11030 shows 99.6% sequence identity with *P. calceolata* a widespread species, whose discovery in 1993 led to the creation of the class Pelagophyceae⁴. Clone OLI11025 is related to the dictyochophytes, which contain phototrophic, phagotrophic and mixotrophic species. Five other clones are affiliated to two highly diverging heterotrophic stramenopile lineages. One group (OLI11026 and OLI11008) clusters with the oomycetes (*Lagenidium*, *Phytophthora* and *Achlya*), whereas the other group (OLI11066, OLI11150 and OLI11006) apparently represents an early heterotrophic divergence. Clone OLI11066 clusters with clone OLI11150, whereas the affinity of clone OLI11006 with the two former clones is not strongly supported by bootstrapping (54%). The relative branching order between early diverging heterotrophic stramenopiles, for example bicoseoids (*Syluania*, *Cafeteria*), labyrinthulids (*Labyrinthuloides minuta*) and thraustochytrids (*Thraustochytrium kinnei*), is still obscure, as shown in previous studies based on 18S rRNA^{15,16}.

Dinoflagellates, like the stramenopiles, contain both autotrophic and heterotrophic taxa. Three clones (OLI11255, OLI11027 and OLI11005) can be included in various dinoflagellate clades, although their exact phylogenetic positions are not clear (bootstrap values <50%). Seven clones (OLI11115, OLI11261, OLI11055, OLI11023, OLI11010, OLI11009 and OLI11012) form a monophyletic lineage (Fig. 1) that includes the parasitic syndiniophycean *Amoebophrya* sp.¹⁷. Additional sequences of syndiniophyceans are probably needed for a more detailed characterization of these clones.

The most intriguing discovery from this work is that of an environmental lineage consisting of six clones (OLI11038,

[†] Present address: Department of Evolutionary Microbiology, University of Nijmegen, NL-6525ED, The Netherlands.

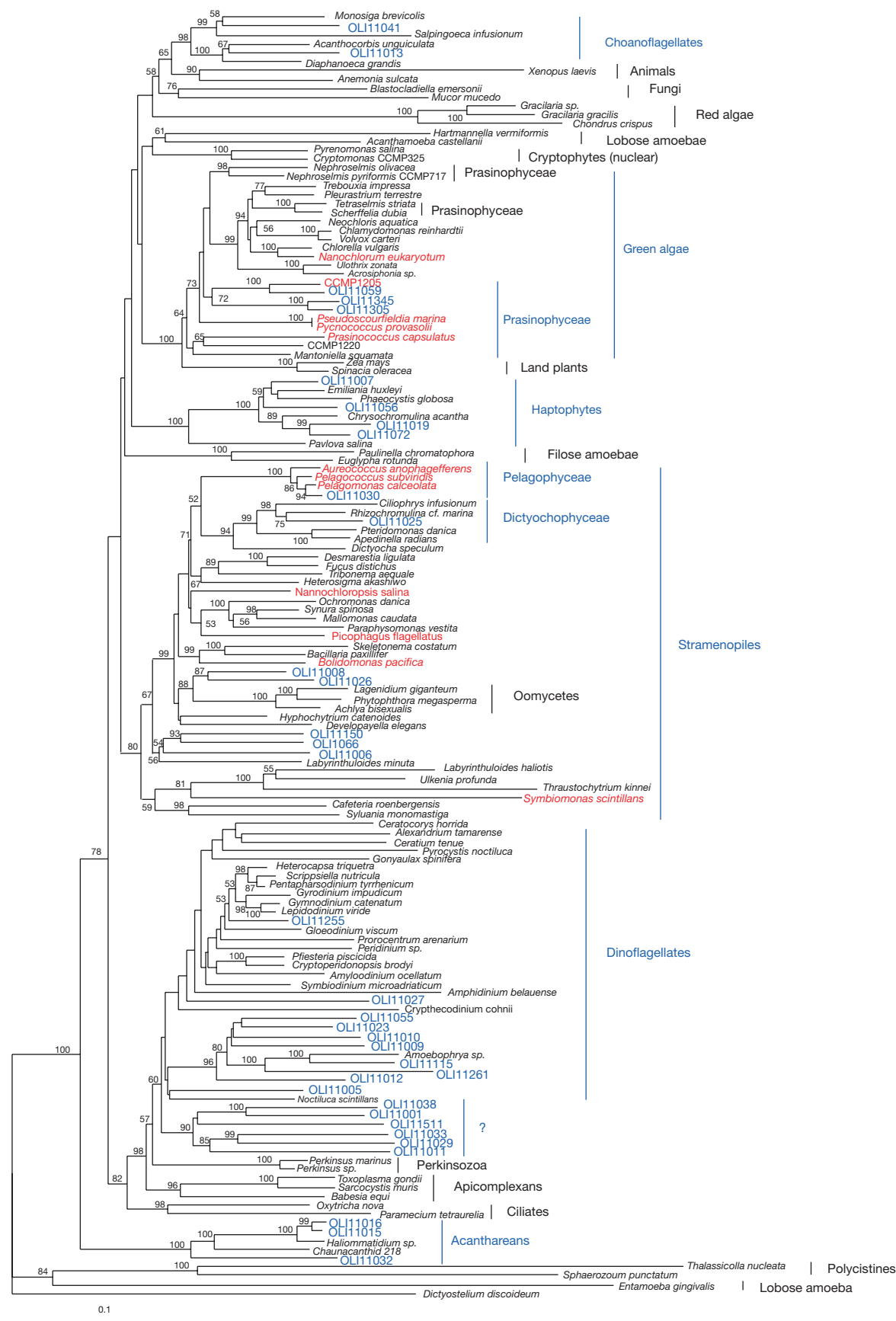


Figure 1 Phylogenetic tree based on nearly complete sequences of the 18S rRNA gene. Sequences in blue beginning with 'OLI11' correspond to those retrieved from the Pacific Ocean (150° W, 11.5° S, 75 m depth). Known marine picoplankton species are indicated

in red. Numbers at nodes represent the bootstrap percentages from 100 replicates. Values below 50% are not shown. The scale bar corresponds to a distance of 10 substitutions per 100 nucleotide positions.

OLI11001, OLI11511, OLI11011, OLI11029 and OLI11033) that cannot be assigned to any known eukaryotic taxonomic group (Fig. 1). It seems to be a close relative of both the extant dinoflagellate lineage and the Perkinsozoa¹⁸, a newly established phylum regrouping the marine parasites *Perkinsus* and *Parvilucifera*, of which the latter infects marine dinoflagellates. The monophyly of the clade combining this environmental lineage and the dinoflagellates was not strongly supported by the bootstrap value (60%). This clade could be related to the heterotrophic species *Oxyrrhis*, currently viewed as a "pre-dinoflagellate"¹⁹. However, because no 18S rDNA sequence is yet available for this taxon, it is not possible to verify this hypothesis.

Among purely heterotrophic protists, acantharians are typical blue-water organisms abundant in temperate and subtropical regions of the world oceans that are important in the global oceanic budget of elements such as strontium and barium²⁰. Clones OLI11015, OLI11016 and OLI11032 cluster with known acantharian sequences (Fig. 1). Clone OLI11015 matches very well (99.2% identity) an arthracanthid sequence from the Atlantic Ocean⁹ and groups together with clones OLI11016 and *Haliommatidium*. Both clone sequences match perfectly the two acantharian probes designed in a previous study¹¹, whereas clone OLI11032 matches only one of these acantharian probes. However, there is a strong bootstrap support for the monophyly of the clade including the clone OLI11032 and the acantharians (100%). The vegetative cells of acantharians are usually much larger than picoplankton. However, the sequences are probably from small swimmers that fall in the size range 3–6 μm (ref. 21).

The last set of sequences (OLI11041 and OLI11013) cluster with the heterotrophic choanoflagellates, which are the closest known relatives to sponges. Clone OLI11041 is related to the naked Codosigidae (*Monosiga*) and OLI11013 to the marine Acanthocidae (*Acanthocorbis*), which display delicate silicified lorica.

Although the diversity of prokaryotes has been the focus of much work with molecular methods during the past decade⁷, quite surprisingly there has been no investigation of eukaryotic diversity in a similar manner⁸. The sequences, which were retrieved from a single sample, collected at the bottom euphotic layer of the Pacific Ocean, suggest that many new eukaryotic lineages remain to be uncovered from the marine environment, especially within the smallest size class. The genuineness of the new lineages found in this study, such as that related to both dinoflagellates and Perkinsozoa, is demonstrated by the fact that they are represented by several different sequences. Nevertheless, this diversity probably constitutes only the tip of the iceberg for three reasons. First, the primers that we used do not target all protists (for example, Chrysophytes or some Apicomplexa). Second, we restricted ourselves to picoplankton (cells smaller than 3 μm), and diversity is probably greater for larger cells²². Last, all sequences originated from a single sample taken at depth in a relatively oligotrophic area: surface waters or more mesotrophic regions probably harbour different organisms. Some species, such as *P. calceolata*, might constitute ubiquitous oceanic 'weeds' in much the same way as the photosynthetic prokaryote *Prochlorococcus*³, whereas others might be restricted to very well defined niches. This might be true of potentially parasitic species, as suggested by the wide diversity of sequences within the clades found between dinoflagellates and Perkinsozoa (Fig. 1).

From the phylogenetic point of view, in many cases the new sequences form distinct clades with regard to organisms known from cultures (for example, OLI11345 and OLI11305 within the Prasinophyceae, or OLI11008 and OLI11026 within the stramenopiles). They might therefore shed light on the evolution of these key protist phyla, as well as for oceanic groups, such as the acantharians, that cannot be maintained in culture. From the ecological point of view, these data suggest that a large fraction of eukaryotic picoplankton is heterotrophic, because 24 of the 35 sequences retrieved

belong to heterotrophic lineages, some of which are probably parasitic (for example, the dinoflagellate lineage containing *Amoebophrya* and the new lineage between dinoflagellates and the Perkinsozoa). This importance of parasitism might require a revision of our current views on the structure of marine microbial food webs, which consider oceanic heterotrophs mainly as phagotrophic predators. As soon as more eukaryotic sequences from a variety of environment become available, it will become possible to design oligonucleotide probes²³ that will enable us to determine the morphology and ecological role of these unknown protists. □

Methods

Picoplankton collection, genomic DNA extraction and PCR amplification of rDNA

Picoplankton was collected at 75 m depth from nutrient-depleted water of the equatorial Pacific Ocean (150°W 11.5°S) during the OLIPAC cruise in November 1994. The water sample was pre-filtered through Nuclepore PE filters 47 mm in diameter with a pore size of 3 μm ; 1 litre of the filtrate was then collected on a glass-fibre filter (Whatman GF/F), which was put into a cryovial filled with lysis buffer (0.75 M sucrose, 400 mM NaCl, 20 mM EDTA, 50 mM Tris-HCl (pH 9.0)), immediately frozen in liquid nitrogen and stored at -80°C. Frozen filters were thawed, then homogenized in cold DNA extraction buffer (10 mM Tris-HCl (pH 8.0); 0.1 M EDTA (pH 8.0); 0.5% SDS). Total genomic DNA was extracted and the 18S rRNA genes were amplified by PCR (30 cycles) with the oligonucleotide primers (5'-ACCTGGTTGATCTGCCAC-3', 5'-TGATCCTTCYGC AGGTTAC-3') complementary to regions of conserved sequences close to the respective 5' (*Escherichia coli* position 7) and 3' (*E. coli* position 1534) termini of the 18S rRNA gene¹⁴. Although these primers recognize most protists, some of the 18S sequences found in the public databases display one or more mismatches. This is true, in particular, for the following groups: Chrysophyceae, some Chlorarachniophyceae, a few Apicomplexa (*Plasmodium*), some Ciliophora (*Euplotes*), and primitive eukaryotes such as Metamonada or Euglenozoa.

Gene cloning and colony hybridization

PCR products were purified with the GeneClean II kit (Qiagen). The ends of the amplified DNA fragments were modified for blunt-ended ligation with T4 kinase and polymerase. The blunt-ended 18S rRNA genes were purified again and inserted into the calf-intestinal-phosphatase-treated *HincII* restriction site of the pGEM3Zf(-) plasmid vector (Promega) and transformed into *E. coli* DH5- α cells. Inserted 18S rRNA genes were screened by colony hybridization with a eukaryote probe (5'-GGGCATCACAGACCTG-3')²⁴ labelled with fluorescein-11-dUTP by using the enhanced chemiluminescence (ECL) 3'-oligo-labelling kit (Amersham). Selected recombinants were grown on a nylon membrane disc, Magna Lift (Micron Separation Inc.) placed on Luria-Bertani agar. A clone containing the plasmid vector was applied to the membrane as a negative control. Lysis of cells, DNA denaturation, and removal of bacterial debris were performed as described¹⁴. The membrane was air-dried and baked at 80°C for 2 h. Prehybridization was performed at 42°C for 30 min in hybridization buffer containing 5 $\times$ SSC, 0.1% (w/v) hybridization buffer component (from the ECL 3'-oligo-labelling and detection kit), 20-fold dilution of liquid block (from the same kit) and 0.02% SDS. Hybridization with the oligonucleotide probes (5–10 ng ml⁻¹) was achieved at 42°C for 12–17 h. Post-hybridization wash conditions included two subsequent washes (15 min each) with 0.1 $\times$ SSC, 0.1% SDS at 42°C. Filters were exposed to a blue-light-sensitive autoradiography film (Hyperfilm-ECL; Amersham) to detect the bound probes, in accordance with the manufacturer's instructions (Amersham).

Sequencing and phylogenetic analysis

A total of 103 clones with a rDNA insert from an 18S rDNA clone library were randomly selected. Plasmid DNAs were purified with the FlexiPrep kit (Pharmacia). The rDNA of each clone was partly sequenced (nucleotide positions 1317–1822 on *Artemia salina*). A sequence identity search with the EMBL databank confirmed that all clones were eukaryotic. Because partial sequences did not provide sufficient information for an unequivocal taxonomic identification, 50 representative variants of the clones were screened on the basis of the partial sequence comparison, and the 18S rRNA genes were completely sequenced. Nucleotide sequences were determined for both strands as described previously¹⁴. Sequence identity searches were conducted with the program FASTA. Secondary structures of environmental sequences fitting the model described in detail elsewhere²⁵ were determined (data available from the authors) by aligning each sequence with the most similar one already present in a database of 18S rRNA sequences²⁶, transposing the secondary structure pattern onto the newly aligned sequence, and making slight adjustments where necessary with the nucleotide sequence editor DCSE²⁷. On the basis of the evaluation by the CHECK_CHIMERA program of the Ribosomal Database Project²⁸, analyses of phylogenetic affiliations for separate sequence domains, and predicted secondary structures, 36 of the 50 complete sequences, which showed no evidence for potential chimaeric gene artefacts, were analysed. One clone showed a sequence identity of 99% to a higher plant, *Musa acuminata*. This sequence and the potential chimaeric gene sequences were not considered further. The environmental sequences were aligned with 111 sequences from a range of eukaryotic evolutionary lineages (accession numbers available from the European small subunit ribosomal RNA database²⁶) by using the CLUSTALX 1.64 program²⁹, with manual refinement based on a

consideration of primary and secondary structures. Sequence positions for which putative homology could not be asserted were excluded. Phylogenetic relationships are inferred by the neighbour-joining method with the PHYLIP package, version 3.5c (ref. 30). Evolutionary distances were calculated with the Kimura two-parameter model (transition:transversion ratio = 2.0). Bootstrap methods provided confidence estimates for tree topology. To avoid potential bias, taxon addition order was randomized.

Received 10 July; accepted 21 November 2000.

- Campbell, L., Nolla, H. A. & Vulot, D. The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. *Limnol. Oceanogr.* **39**, 954–961 (1994).
- Li, W. K. W. Primary productivity of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol. Oceanogr.* **39**, 169–175 (1994).
- Partensky, F., Hess, W. R. & Vulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microb. Mol. Biol. Rev.* **63**, 106–127 (1999).
- Andersen, R. A., Saunders, G. W., Paskind, M. P. & Sexton, J. Ultrastructure and 18S rRNA gene sequence for *Pelagomonas calceolata* gen. and sp. nov. and the description of a new algal class, the Pelagophyceae classis nov. *J. Phycol.* **29**, 701–715 (1993).
- Guillou, L. et al. *Bolidomonas*: a new genus with two species belonging to a new algal class, the Bolidophyceae (Heterokonta). *J. Phycol.* **35**, 368–381 (1999).
- Potter, D., Lajeunesse, T. C., Saunders, G. W. & Andersen, R. A. Convergent evolution masks extensive biodiversity among marine coccoid picoplankton. *Biodiversity Conserv.* **6**, 99–107 (1997).
- Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60–63 (1990).
- Rappé, M. S., Suzuki, M. T., Vergin, K. L. & Giovannoni, S. J. Phylogenetic diversity of ultraplankton plastid small-subunit rRNA genes recovered in environmental nucleic acid samples from the Pacific and Atlantic coasts of the United States. *Appl. Environ. Microbiol.* **64**, 294–303 (1998).
- Zettler, L. A. A. & Caron, D. A. New insights into the phylogeny of the Acantharea based on SSU rRNA gene sequencing. *Eur. J. Protistol.* **36**, 34–39 (2000).
- Melkonian, M. Systematics and evolution of the algae. *Prog. Bot.* **57**, 281–311 (1996).
- Zettler, L. A., Sogin, M. L. & Caron, D. A. Phylogenetic relationships between the Acantharea, and the Polycystinea: a molecular perspective on Haeckel's Radiolaria. *Proc. Natl Acad. Sci. USA* **94**, 11411–11416 (1997).
- Van de Peer, Y. & De Wachter, R. Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *J. Mol. Evol.* **45**, 619–630 (1997).
- Venrick, E. L. Phytoplankton species structure in the central North Pacific, 1973–1996: variability and persistence. *J. Plankton Res.* **21**, 1029–1042 (1999).
- Moon-van der Staay, S. Y. et al. Abundance and diversity of prymnesiophytes in the picoplankton community from the equatorial Pacific Ocean inferred from 18S rDNA sequences. *Limnol. Oceanogr.* **45**, 98–109 (2000).
- Leipe, D. D. et al. 16S-like rDNA sequences from *Developayella elegans*, *Labyrinthuloides halitidis*, and *Proteromonas lacertae* confirm that the stramenopiles are a primarily heterotrophic group. *Eur. J. Protistol.* **32**, 449–458 (1996).
- Guillou, L., Chrétiennot-Dinet, M. J., Boulben, S., Moon-van der Staay, S. Y. & Vulot, D. *Symbiomonas scintillans* gen. et sp. nov. and *Picophagus flagellatus* gen. et sp. nov. (Heterokonta): two new heterotrophic flagellates of picoplanktonic size. *Protist* **150**, 383–398 (1999).
- Gundersen, J. H., Goss, S. H. & Coats, D. W. The phylogenetic position of *Amoeboophrya* sp. infecting *Gymnodinium sanguineum*. *J. Euk. Microbiol.* **46**, 194–197 (1999).
- Noren, F., Moestrup, O. & Rehnstam-Holm, A. S. *Parvulicifera infectans* Noren et Moestrup gen. et sp. nov. (Perkinsozoa phylum nov.): a parasitic flagellate capable of killing toxic microalgae. *Eur. J. Protistol.* **35**, 233–254 (1999).
- Fensome, R. et al. *Classification of Living and Fossil Dinoflagellates* (Micropaleontology Press, Hanover, 1993).
- Bernstein, R. E., Byrne, R. H. & Schijf, J. Acantharians: a missing link in the oceanic biogeochemistry of barium. *Deep-Sea Res.* **45**, 491–505 (1998).
- Hollande, A. Données ultrastructurales sur les isopores des radiolaires. *Protistologica* **10**, 567–572 (1974).
- May, R. How many species are there on earth? *Science* **241**, 1441–1449 (1988).
- Amann, R. & Kuhl, M. *In situ* methods for assessment of microorganisms and their activities. *Curr. Opin. Microbiol.* **1**, 352–358 (1998).
- Giovannoni, S. J., DeLong, E. F., Olsen, G. J. & Pace, N. R. Phylogenetic group-specific oligodeoxynucleotide probes for identification of single microbial cells. *J. Bacteriol.* **170**, 720–726 (1988); Erratum, *J. Bacteriol.* **170**, 2418 (1988).
- Van de Peer, Y., Jansen, J., De Rijk, P. & De Wachter, R. Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res.* **25**, 111–116 (1997).
- Van de Peer, Y., De Rijk, P., Wuyts, J., Winkelmans, T. & De Wachter, R. The European small subunit ribosomal RNA database. *Nucleic Acids Res.* **28**, 175–176 (2000).
- De Rijk, P. & De Wachter, R. DCE, an interactive tool for sequence alignment and secondary structure research. *Comput. Appl. Biosci.* **9**, 735–740 (1993).
- Maidak, B. L. et al. The Ribosomal Database Project. *Nucleic Acids Res.* **22**, 3485–3487 (1994).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Felsenstein, J. PHYLIP (Phylogeny Inference Package), version 3.5c (Department of Genetics, Univ. Washington, Seattle, 1993).

Acknowledgements

S.Y.M.-v.d.S. was supported by a 'Poste Rouge' from CNRS. This work was also funded by the European Framework Programme 5 PICODIV and the French programmes ACC-SV7 Biosystematics network and PROOF PROSOPE.

Correspondence and requests for materials should be addressed to D.V. (e-mail: vulot@sb-roscoff.fr). The sequences reported in this paper have been deposited in GenBank under accession numbers AJ402325–AJ402359.

Parallel adaptive radiations in two major clades of placental mammals

Ole Madsen^{*†}, Mark Scally^{†‡§}, Christophe J. Douady^{‡§}, Diana J. Kao[‡], Ronald W. DeBry^{||}, Ronald Adkins[¶], Heather M. Amrine^{‡#}, Michael J. Stanhope^{§*}, Wilfried W. de Jong^{***} & Mark S. Springer^{‡#}

^{*} Department of Biochemistry, University of Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands

[‡] Department of Biology, University of California, Riverside, California 92521, USA

[§] Queen's University of Belfast, Biology and Biochemistry, 97 Lisburn Road, Belfast, BT9 7BL, United Kingdom

^{||} Department of Biological Sciences, Box 210006, University of Cincinnati, Cincinnati, Ohio 45221 USA

[¶] Biology Department, University of Massachusetts, Amherst, Massachusetts, 01003 USA

[#] Graduate Group in Genetics, University of California, Riverside, California 92521 USA

^{*} Bioinformatics, SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, UP1345, Collegeville, Pennsylvania 19426 USA

^{***} Institute for Biodiversity and Ecosystem Dynamics, 1090 GT Amsterdam, The Netherlands

[†] These authors contributed equally to this work

Higher level relationships among placental mammals, as well as the historical biogeography and morphological diversification of this group, remain unclear^{1–3}. Here we analyse independent molecular data sets, having aligned lengths of DNA of 5,708 and 2,947 base pairs, respectively, for all orders of placental mammals. Phylogenetic analyses resolve placental orders into four groups: Xenarthra, Afrotheria, Laurasiatheria, and Euarchonta plus Glires. The first three groups are consistently monophyletic with different methods of analysis. Euarchonta plus Glires is monophyletic or paraphyletic depending on the phylogenetic method. A unique nine-base-pair deletion in exon 11 of the *BRCA1* gene provides additional support for the monophyly of Afrotheria, which includes proboscideans, sirenians, hyraxes, tubulidentates, macroscelideans, chrysochlorids and tenrecids. Laurasiatheria contains cetartiodactyls, perissodactyls, carnivores, pangolins, bats and eulipotyphlan insectivores. Parallel adaptive radiations have occurred within Laurasiatheria and Afrotheria. In each group, there are aquatic, ungulate and insectivore-like forms.

The combination of fossil and anatomical data has suggested moderately well-resolved phylogenetic trees for the 18 extant orders of placental mammals^{1,2}. DNA sequences have provided consistent support for only a few of the proposed superordinal groups, notably for Paenungulata (elephants, sea cows and hyraxes) and Cetartiodactyla (artiodactyls and whales)³, and have rejected some traditional clades such as Archonta (primates, tree shrews, flying lemurs and bats)⁴. Molecular data have also suggested new sets of relationships; Afrotheria, a group that includes paenungulates, aardvarks, elephant shrews, golden moles and tenrecs, is supported by both mitochondrial ribosomal RNA and nuclear protein-coding genes^{5,6}. These same sequences suggest that lipotyphlan insectivores are paraphyletic or polyphyletic.

Increased resolving power may result from concatenations of individual genes⁷. We concatenated DNA sequences for mitochondrial RNA genes and three nuclear genes (A2AB, IRBP, vWF), including 16 new sequences (see Methods). This data set includes 26 placental taxa, representative of all eutherian orders, and two marsupial outgroups (see Fig. 1A). In addition, golden mole and tenrec, belonging to the insectivore families Chrysochloridae and Tenrecidae, respectively, are represented by sequences for all genes

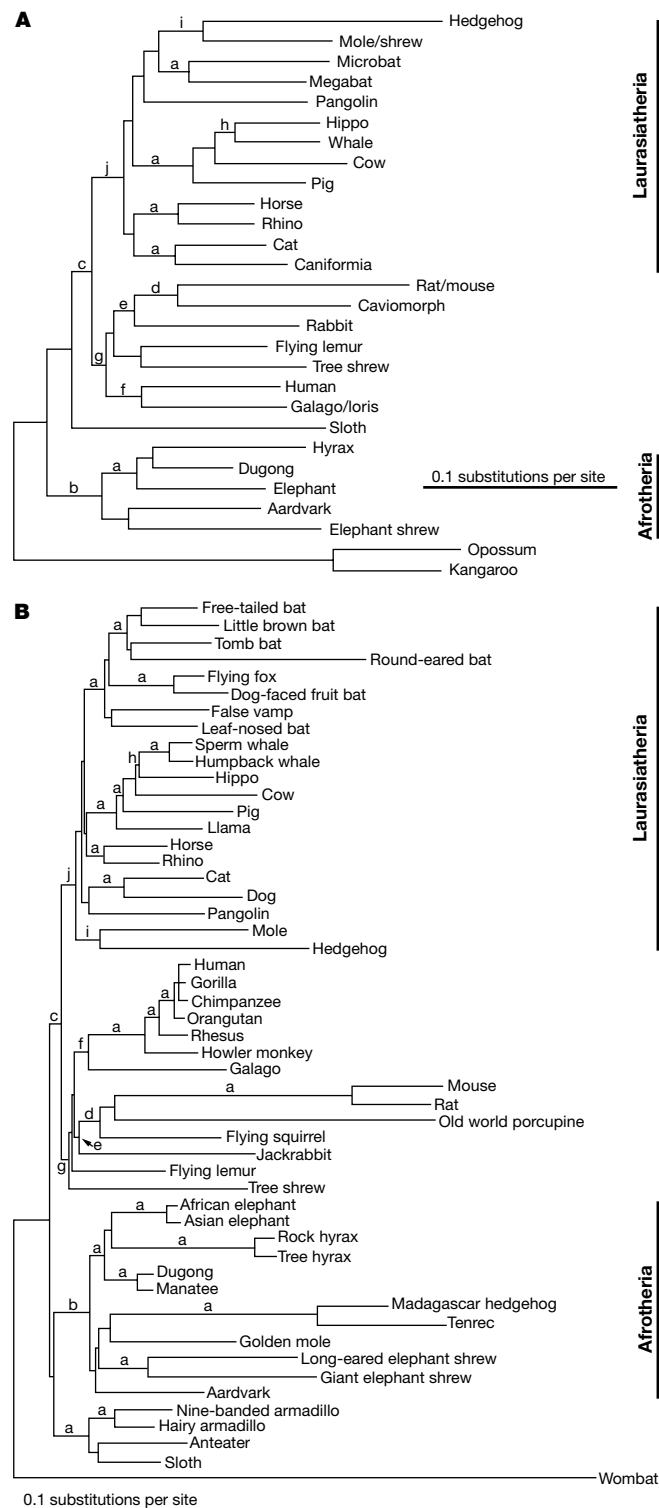


Figure 1 Rooted maximum-likelihood trees. **A**, The 5,708-bp data set. **B**, The 2,947-bp data set. For chimaeric sequences in **A**, the combined species or the higher taxonomic unit are indicated. Branch lengths are proportional to the amount of sequence change (scale bar, 10% sequence divergence). Bootstrap support values for nodes b–j with different phylogenetic methods are shown in Table 1; nodes labelled a are supported at or above 90% by all methods. Afrotheria also includes golden moles and tenrecs, which are not represented by IRBP sequences, and therefore not shown in **A**; in analyses with the 5,708-bp data set that included golden mole and tenrec, bootstrap support for Afrotheria ranged from 88% to 100%. For the 5,708-bp data set, the 12 statistically acceptable

branches for the location of the root, in decreasing order of likelihood score, are branch b (as shown), sloth, rabbit, branch leading to rat/mouse, branch c, branch d, branch e, caviomorph, branch j, tree shrew, branch leading to flying lemur + tree shrew, and galago/loris; other branches were rejected. For the 2,947-bp data set, the statistically acceptable options, in decreasing order of likelihood score, were branch c (as shown), branch b, and the base of Xenarthra. The name Laurasiatheria ("from the area of Laurasia or Europe + Asia + North America")¹⁸ was suggested for this clade, but no support in the form of bootstrap analyses and/or statistical tests was provided.

except IRBP; inclusion of these taxa in a subset of analyses was necessary to investigate the proposed diphyly of the insectivores and to confirm the naturalness of Afrotheria. In all, 5,708 aligned nucleotide positions were used in phylogenetic analyses.

Hypothesis testing with independent data sets is fundamental in phylogenetics. To establish a data set that was independent of the 5,708-bp concatenation, we sequenced roughly 3 kilobases (kb) of exon 11 of the single-copy breast and ovarian cancer susceptibility gene 1 (*BRCA1*) for 33 taxa and combined these with 19 GenBank sequences (see Methods). The aligned *BRCA1* data set includes representatives of all placental orders and one marsupial outgroup (for names, see Fig. 1B). Finally, we combined the 5,708-bp and 2,947-bp data sets (total 8,655 bp) for 26 overlapping placentals and one marsupial.

Phylogenetic trees (maximum likelihood) for the 5,708-, 2,947- and 8,655-bp data sets are shown in Fig. 1A and B, and Supplementary Information, respectively. Bootstrap support values for different phylogenetic methods are given in Table 1 for the 5,708- and 2,947-bp data sets, and in Supplementary Information for the 8,655-bp data set. In spite of the differences in taxonomic sampling, as well as some conspicuously long branches on the *BRCA1* tree, placental orders are resolved into the following four groups on all three maximum-likelihood trees: Xenarthra, Afrotheria, Laurasiatheria, and Glires (rodents and lagomorphs) + Euarchonta (flying lemurs, tree shrews and primates). Xenarthra, Afrotheria, and Laurasiatheria are robust with different phylogenetic methods (Fig. 1B; Table 1; Supplementary Information). Afrotheria is also supported by a 9-bp deletion that is shared by all 12 afrotherians that were included in our study (Fig. 2). The absence of this deletion in other placentals, as well as in marsupials, indicates that this feature may be a shared derived character of Afrotheria. Despite the absence of any morphological evidence for Afrotheria^{2,8}, it now becomes increasingly difficult to explain this hypothesis except in the context of shared common ancestry. Within Afrotheria, there is strong support for Paenungulata. Our analyses provide the first robust support for the Laurasiatheria clade. Within this clade, there is strong support for Cetartiodactyla and the monophyly of eulipotyphlan insectivores (hedgehogs, shrews and moles).

Support for Euarchonta + Glires was inconsistent, largely because the position of the root was sensitive to the phylogenetic method that was used (Table 1; Supplementary Information). Maximum-likelihood analyses with the *BRCA1* data set provided 100% bootstrap support for the monophyly of Glires + Euarchonta, whereas parsimony provided only 18% support. The shortest parsimony trees (two at 10,867 steps; Supplementary Information) for *BRCA1* root on Old World porcupine, rendering Euarchonta + Glires, Glires and Rodentia paraphyletic; 35 additional steps are required to root in the same location as the likelihood tree shown in Fig. 1B. Among the individual bootstrap replicate trees for parsimony, rootings occurred on long, topologically separated branches as

follows: Old World porcupine (69%); rat-mouse (11%); Madagascar hedgehog-tenrec (11%); tree shrew (4%); hedgehog (3%); and round-eared bat (2%). Maximum likelihood did not root on these long branches. Instead, there was 100% bootstrap support for the monophyly of Laurasiatheria + Euarchonta + Glires (branch c), and individual bootstrap replicate trees, with or without an allowance for rate-heterogeneity, rooted on branch b (base of Afrotheria), branch c or at the base of Xenarthra; these three branches are topologically adjacent, which suggests a strong affinity of the root for this general region of the tree. Furthermore, statistical tests with likelihood reject rooting the *BRCA1* tree on all branches except for these three. Finally, Monte Carlo simulations mirrored our results for the actual *BRCA1* data set and showed that parsimony will root on long branches and fail to recover the correct root given the topology depicted in Fig. 1B (Supplementary Information).

Twelve roots are statistically acceptable on the likelihood tree for the 5,708-bp data set (Fig. 1A). The three possible *BRCA1* roots coincide with three of the five best roots on the 5,708-bp data set tree. These are also the three best roots for the 8,655-bp data set. Although several roots in the Euarchonta + Glires group could not be rejected with the 5,708- and 8655-bp data sets, we do not favour these roots. First, there are numerous morphological synapomorphies for both rodent monophyly⁹ and Glires². Second, other molecular results also favour rodent monophyly and even the monophyly of Glires, when phylogenetic methods are used that address problems such as nonstationarity and long-branch attraction^{10,11}.

Afrotheria and Xenarthra have observed first occurrences in the fossil record on Southern Hemisphere landmasses—Africa in the case of Afrotheria^{5,6} and South America in the case of Xenarthra¹². First occurrences in the fossil record do not permit unequivocal assignment of continental origin. Nevertheless, there is a strong possibility that Afrotheria and Xenarthra have Gondwanan origins. Of the three roots (base of Afrotheria, branch b; base of Laurasiatheria + Euarchonta + Glires, branch c; base of Xenarthra) that are statistically acceptable on all three trees, rooting on the xenarthran branch is the only hypothesis that is consistent with the morphology-based Epitheria hypothesis². At the same time, rooting on the Xenarthra branch results in paraphyly of a possible Gondwanan group (i.e., Xenarthra + Afrotheria) at the base of the placental tree. Rooting on branch b also results in paraphyly of this Southern Hemisphere group. The suggestion that a Gondwanan assemblage is paraphyletic at the base of the tree contrasts with traditional views on the evolution of placental mammals, which place their most recent common ancestor in the Northern Hemisphere¹³. Considering the debate about the timing of the principal placental divergences^{14,15}, the paucity of Cretaceous mammal fossils from the Southern Hemisphere¹⁵, and the discoveries of tribosphenic fossils in the Southern Hemisphere, including a possible placental¹⁶, a Gondwanan origin for extant placentals should not be excluded. A Southern Hemisphere ancestry for

Table 1 Bootstrap support percentages for branches b–j in Fig. 1 with different phylogenetic methods

Node	Parsimony				Distance methods						Maximum likelihood			
	Unweighted		Transversion		NJ-WAVE		ML-GTR		Logdet		No rate heterogeneity		Rate heterogeneity	
	5,708	2,947	5,708	2,947	5,708	2,947	5,708	2,947	5,708	2,947	5,708	2,947	5,708	2,947
b	100	89	100	83	100	96	100	76	100	94	100	100	100	100
c	53	11	62	29	51	56	19	40	30	39	45	100	64	100
d	79	22	74	30	41	69	99	48	100	61	80	99	94	99
e	53	3	4	26	5	34	42	17	54	15	41	67	59	68
f	97	88	71	83	98	89	99	84	99	85	99	99	96	98
g	48	18	58	30	6	49	5	35	10	34	44	100	66	100
h	93	13	72	42	100	85	100	88	100	88	98	36	96	44
i	95	95	94	100	100	100	76	99	74	99	100	100	96	100
j	96	93	87	97	76	97	81	89	78	98	99	100	99	100

In the first row, 5,708 and 2,947 refer to the 5,708-bp and 2,947-bp data sets, respectively. In maximum-likelihood analyses that did not allow for rate heterogeneity, transition/transversion ratios were set at 1.81 and 2.22 for the 5,708- and 2,947-bp data sets, respectively. In maximum-likelihood analyses that allowed for rate heterogeneity, transition/transversion ratios were set at 1.89 and 2.24 for the 5,708- and 2,947-bp data sets, respectively. ML-GTR, maximum likelihood with general time reversible model; WAVE, weighted average distances (see Methods).

Our molecular results support morphology-based hypotheses such as Paenungulata. Other hypotheses, such as Ungulata (cetartiodactyls, paenungulates, perissodactyls and tubulidentates), Altungulata (paenungulates and perissodactyls), Anagalida (Glires and elephant shrews), Lipotyphla, Archonta, Volitantia (bats and flying lemurs), Artiodactyla and Edentata (xenarthrans and pholidotes) are rejected with statistical tests (Supplementary Information) in favour of Afrotheria and Laurasiatheria. If hypotheses such as Afrotheria and Laurasiatheria are correct, then morphology has failed to recover some of the most fundamental clades in all of

Matthew²⁰ suggested that “insectivores” were the central stock from which other eutherian orders originated, and there is abundant evidence for insectivore-like mammals in the fossil record²¹. Even today, living lipotyphlan insectivores and fossil insectivores such as the Cretaceous palaeoryctids are often regarded as among the more primitive eutherian mammals^{1,2}. Easteal²² suggested that primitive placentals from the Cretaceous may have diversified phylogenetically before they diverged morphologically and acquired the diagnostic features of ordinal level crown-group clades. The present results, which deploy living lipotyphlan insectivores into Laurasiatheria and Afrotheria, respectively, are consistent with the view that insectivore-like forms were the central stock from which other groups originated and further suggest that different insectivore-like lineages have independently given rise to diverse eutherian taxa in different geographic venues. The finding that parallel adaptive radiations have occurred within Eutheria also helps to explain the difficulties that have accompanied efforts to elucidate interordinal relationships based on morphology. □

Sequences were aligned using CLUSTAL W²³. Ambiguous regions of the 12S rRNA/tRNA valine-16S rRNA alignment were excluded from phylogenetic analyses as were a glutamic

Figure 2 Positions 353–383 of the BRCA1 nucleotide alignment for representative taxa to show the 9-bp deletion that occurs in all 12 afrotherians but not in 39 other placental mammals or in the 3 marsupials that were investigated.

acid repeat region of the A2AB gene and a 21-bp region of the *BRCA1* gene that is repeated up to four times. The aligned data sets were the following lengths: A2AB (1,164 bp); IRBP (1,292 bp); vWF (1,251 bp); 12S rRNA/tRNA valine-16S rRNA (2,001 bp); and *BRCA1* (2,947 bp). Alignments for the concatenated (A2AB + IRBP + vWF + rRNA) and *BRCA1* data sets are available in the Supplementary Information. Phylogenetic analyses included unweighted and transversion parsimony, minimum evolution (5,708-bp data set) or neighbour joining (*BRCA1* data set) with logdet and maximum likelihood-GTR²⁴ distances, neighbour joining with weighted average (WAVE) maximum likelihood distances²⁵, and maximum likelihood under the HKY85 (ref. 24) model of sequence evolution. Gaps were coded as missing in parsimony analyses. Maximum-likelihood estimates of relative rates and transition to transversion ratios were obtained from maximum parsimony trees and used in subsequent maximum-likelihood analyses and in calculating weighted-average maximum-likelihood distances. Bootstrap support values are based on 500 replications except for maximum likelihood (100 replications). Maximum-likelihood bootstrap analyses with the *BRCA1* data set used the following backbone constraint, where taxon numbers correspond to the ordering of taxa (top to bottom) in Fig. 1B: (((1–4), (5, 6), (7, 8)), ((9, 10), 11–14), (15, 16), (17, 18), 19–21, (((22–24), 25), 26), 27), 28, (29, 30), 31–35, ((36, 37), (38, 39), (40, 41)), ((42, 43), 44), (45, 46), 47, ((48, 49), 50, 51)). Kishino and Hasegawa tests²⁴ were used to examine *a priori* hypotheses and to examine statistically acceptable root locations. In the latter case, we obtained the best unrooted likelihood tree for each data set and then evaluated all possible root positions. All phylogenetic analyses and statistical tests were performed with PAUP 4.0b2 (ref. 26), except for neighbour-joining with weighted average distances, where analyses were performed with PHYLIP 3.572 (J. Felsenstein) and WAVEBOOT (D. King and C. Krajewski). Maximum-likelihood analyses with rate partitions allowed the following eight rate partitions with the 5,708-bp data set: third positions of each nuclear gene; first + second positions of each nuclear gene; RNA stems; and RNA loops. Two rate partitions, corresponding to first + second and third codon positions, respectively, were used with the *BRCA1* data set. NJ-WAVE analyses used a weighted-average distance approach²⁵ with the eight partitions indicated above for the 5,708-bp data set and two partitions (first + second codon positions; third codon positions) for the 2,947-bp data set; each partition was allowed its own rate, base composition, and transition to transversion ratio. Monte Carlo simulations were performed with Seq-Gen 1.1 (ref. 27) (Supplementary Information). Molecular dates were estimated using QDATE²⁸ (Supplementary Information).

Received 30 August; accepted 10 October 2000.

- Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356**, 121–125 (1992).
- Shoshani, J. & McKenna, M. C. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. *Mol. Phylogenet. Evol.* **9**, 572–584 (1998).
- de Jong, W. W. Molecules remodel the mammalian tree. *Trends Ecol. Evol.* **13**, 270–275 (1998).
- Springer, M. S., Burk, A., Kavanagh, J. R., Waddell, V. G. & Stanhope, M. J. The interphotoreceptor retinoid binding protein gene in therian mammals: implications for higher level relationships and evidence for loss of function in the marsupial mole. *Proc. Natl Acad. Sci. USA* **94**, 13754–13759 (1997).
- Springer, M. S. *et al.* Endemic African mammals shake the phylogenetic tree. *Nature* **388**, 61–64 (1997).
- Stanhope, M. J. *et al.* Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl Acad. Sci. USA* **95**, 9967–9972 (1998).
- Teeling, E. C. *et al.* Molecular evidence regarding the origin of echolocation and flight in bats. *Nature* **403**, 188–192 (2000).
- Asher, R. J. A morphological basis for assessing the phylogeny of the 'Tenrecidae' (Mammalia, Lipotyphla). *Cladistics* **15**, 231–252 (1999).
- Lockett, W. P. & Hartenberger, J.-L. Monophyly or polyphyly of the order Rodentia: Possible conflict between morphological and molecular interpretations. *J. Mammal. Evol.* **1**, 127–147 (1993).
- Waddell, P. J., Cao, Y., Hauf, J. & Hasegawa, M. Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid-invariant sites-logdet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo and elephant. *Syst. Biol.* **48**, 31–53 (1999).
- Penny, D., Masegawa, M., Waddell, P. J. & Hendy, M. D. Mammalian evolution: Timing and implications from using the logdeterminant transform for proteins of differing amino acid composition. *Syst. Biol.* **48**, 76–93 (1999).
- McKenna, M. C. & Bell, S. K. *Classification of Mammals Above the Species Level* (Columbia Univ. Press, New York, 1997).
- Rainger, R. *Agenda for Antiquity: Henry Fairfield Osborn and Vertebrate Paleontology at the American Museum of Natural History, 1890–1935* (Univ. Alabama Press, Tuscaloosa, Alabama, 1991).
- Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
- Foot, M., Hunter, J. P., Janis, C. M. & Sepkoski, J. J. Jr Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. *Science* **283**, 1310–1314 (1999).
- Rich, T. H. *et al.* A tribosphenic mammal from the Mesozoic of Australia. *Science* **278**, 1438–1442 (1997).
- Mouchaty, S. K., Gullberg, A., Janke, A. & Arnason, U. The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. *Mol. Biol. Evol.* **17**, 60–67 (2000).
- Waddell, P. J., Okada, N. & Hasegawa, M. Towards resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48**, 1–5 (1999).
- Rasmussen, D. T. In *The Evolution of Perissodactyls* (eds Prothero, D. R. & Schoch, R. M.) 57–78 (Oxford Univ. Press, Oxford, 1989).
- Matthew, W. D. The Carnivora and Insectivora of the Bridger basin, middle Eocene. *Mem. Am. Nat. Hist.* **9**, 291–567 (1909).
- Novacek, M. J. The skull of lepidictid insectivores and the higher-level classification of eutherian mammals. *Bull. Am. Mus. Nat. Hist.* **183**, 1–111 (1986).
- Easteal, S. Molecular evidence for the early divergence of placental mammals. *BioEssays* **21**, 1052–1058 (1999).

- Thompson, J. D., Higgins, G. D. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Swofford, D. L., Olsen, G. P., Waddell, P. J. & Hillis, D. M. In *Molecular Systematics* (eds Hillis, D. M., Moritz, C. & Mable, B. K.) 407–492 (Sinauer, Sunderland, Massachusetts, 1996).
- Krajewski, C., Blacket, M., Buckley, L. & Westerman, M. A multigene assessment of phylogenetic relationships within the dasyurid marsupial subfamily Sminthopsinae. *Mol. Phylogenet. Evol.* **8**, 236–248 (1997).
- Swofford, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods) Version 4, (Sinauer, Sunderland, Massachusetts, 1998).
- Rambaut, A. & Grassly, N. C. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13**, 303–306 (1997).
- Rambaut, A. & Bromham, L. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15**, 442–448 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank F. Catzeflis for tissue samples. This work was supported by the NSF (M.S.S.) and the TMR program of the European Commission (W.W.d.J.; M.J.S.).

Correspondence and requests for materials should be addressed to M.S.S. (e-mail: mark.springer@ucr.edu).

Molecular phylogenetics and the origins of placental mammals

William J. Murphy^{*†}, Eduardo Eizirik^{*‡†}, Warren E. Johnson^{*},
Ya Ping Zhang[§], Oliver A. Ryder^{||} & Stephen J. O'Brien^{*}

^{*} Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland 21702, USA

[§] Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

^{||} Center for Reproduction of Endangered Species, Zoological Society of San Diego, San Diego, California 92112, USA

[‡] Department of Biology, University of Maryland, College Park, Maryland 20742, USA

[†] These authors contributed equally to this work

The precise hierarchy of ancient divergence events that led to the present assemblage of modern placental mammals has been an area of controversy among morphologists, palaeontologists and molecular evolutionists. Here we address the potential weaknesses of limited character and taxon sampling in a comprehensive molecular phylogenetic analysis of 64 species sampled across all extant orders of placental mammals. We examined sequence variation in 18 homologous gene segments (including nearly 10,000 base pairs) that were selected for maximal phylogenetic informativeness in resolving the hierarchy of early mammalian divergence. Phylogenetic analyses identify four primary superordinal clades: (I) Afrotheria (elephants, manatees, hyraxes, tenrecs, armadillo and elephant shrews); (II) Xenarthra (sloths, anteaters and armadillos); (III) Glires (rodents and lagomorphs), as a sister taxon to primates, flying lemurs and tree shrews; and (IV) the remaining orders of placental mammals (cetaceans, artiodactyls, perissodactyls, carnivores, pangolins, bats and core insectivores). Our results provide new insight into the pattern of the early placental mammal radiation.

The panoply of morphological, ecological and genomic diversity among extant mammals offers considerable potential for studies of speciation, adaptation, molecular evolution, genome organization and biogeography^{1–3}. Studies on morphology and both mitochondrial and nuclear genes have revealed several higher level phylogenetic associations^{2,4–7}, but a full resolution of the earliest placental

divergences has yet to be accomplished. We obtained sequences from segments of 15 nuclear and three mitochondrial genes (9,779 base pairs (bp)) in 64 placental and two marsupial species (Table 1). Phylogenetic analyses of the concatenated data set using maximum parsimony, maximum likelihood and distance based (neighbour joining) methods all converged on a nearly identical, well supported topology defining four principal eutherian lineages (Fig. 1). The results affirm monophyly of traditional placental orders (except Artiodactyla and Insectivora), and also support some previously proposed, as well as new, superordinal clades.

Our analyses provide further support for the superordinal clade Afrotheria^{4,5,8} (Fig. 1, clade I), and place it as the most basal placental divergence. Within Afrotheria, our data support a consistently strong clustering of Paenungulata⁹ (Hyracoidea, Sirenia and Proboscidea) and a nested Tethytheria¹⁰ (Sirenia and Proboscidea). The next basal mammalian lineage in our trees was the Neotropical order Xenarthra (Fig. 1, clade II), whose monophyly was strongly supported in all analyses. Within Xenarthra, our data consistently united sloths (*Choloepus* spp.) and anteaters (*Myrmecophaga*, *Tamandua*) in the proposed suborder Pilosa⁶, which was in turn the sister group to armadillos. Although all of our trees showed the placental root to be either between Afrotheria and all other placentals (neighbour joining and maximum likelihood), or within Afrotheria (maximum parsimony; on the branches leading to elephant shrews or the tenrec), we could not statistically reject a scenario with Afrotheria and Xenarthra as sister groups, separated from all other eutherians by a basal split ($P \geq 0.07$ in Kishino–Hasegawa (parsimony and likelihood) and Templeton tests). Furthermore, we could not statistically reject a scenario showing Xenarthra basal to all other placentals ($P \geq 0.50$ in Kishino–Hasegawa (parsimony and likelihood) and Templeton tests), suggesting that additional data collection will be required to resolve the placental root with a high level of confidence.

After these two basal divergences are two internal superordinal clades of eutherians (Fig. 1, clades III and IV), which were found as sister groups in all of our analyses. Clade III represents a novel group uniting rodents, lagomorphs, primates, tree shrews (Scandentia) and flying lemurs (Dermoptera). This is in contrast to molecular studies that show support for rodents being basal eutherians^{7,11–13}, which was possibly influenced by incomplete taxon sampling and extreme rate acceleration in this group. Although bootstrap support for this clade is not as high as for the other principal clades (neighbour joining = 93%, maximum likelihood = 85%, maximum parsimony = 64% and weighted parsimony = 73%), it is consistently resolved in all of our analyses. Our increased taxon sampling provides robust support for the monophyly of rodents, as well as

their sister-group relationship with lagomorphs (cohort Glires^{9,14}), a grouping that had been well established on morphological grounds² but has been contradicted or unresolved with molecular approaches^{13,15,16}.

The cohort Glires is, in turn, the sister group of primates, flying lemurs and tree shrews. Within primates, lemur (Strepsirhini) and tarsier (Tarsiiformes) were found to be sister taxa (bootstrap $\geq 80\%$) that were separated from anthropoids by a deep divergence. This contrasts with the widely held view that tarsiers are more closely related to anthropoids than to Strepsirhini¹⁷. Furthermore, although maximum parsimony and maximum likelihood analyses consistently supported the monophyly of Primates and the sister-group relationship between Dermoptera and Scandentia, distance-based trees suggested primate paraphyly by placing flying lemurs as the sister group to anthropoids, and by placing tree shrews basal among the three orders (Fig. 1). Kishino–Hasegawa (parsimony and likelihood) and Templeton tests failed to reject either of these hypotheses ($P > 0.1$ in all cases), which indicates that further sampling within Dermoptera and Scandentia may be required to fully resolve the deep, contemporaneous divergence among these three orders.

Seven remaining placental orders—cetaceans, artiodactyls, perisodactyls, carnivores, pangolins, bats, and core insectivores (hedgehogs, shrews and moles)—comprise a fourth superordinal lineage (Fig. 1, clade IV), a grouping that has received support from whole-mitochondrial genome sequence analysis (although no pangolin was sampled)⁷. Our results affirm the monophyly of Chiroptera (megabats plus microbats), which has been a topic of debate for over a decade¹⁸. These data also reject ($P < 0.0001$ for all tests) a direct relationship of bats with primates, flying lemurs and tree shrews, which were classically proposed to form the cohort Archonta¹⁴. Our data support the polyphyly of Insectivora^{4,5,8}, with shrews, moles and hedgehogs forming a monophyletic group provisionally termed 'Eulipotyphla'¹⁹, whereas tenrecs clearly cluster within the superordinal clade Afrotheria. We find strong support for the traditional view of placing hedgehogs in a clade with shrews and moles (bootstrap $\geq 99\%$), which is in contrast to the basal position within eutherians observed with mitochondrial DNA (mtDNA) genomes²⁰.

The inter-ordinal relationships within clade IV were not fully resolved with high confidence, which is possibly due to the very rapid diversification of this group. However, our parsimony and maximum likelihood analyses agree with recent mtDNA data⁷, suggesting a sister-group relationship between Chiroptera and Eulipotyphla at the base of this principal clade (see Supplementary Information). The pangolin (Pholidota) was the sister group to carnivores in combined (and many of the single-gene) analyses, a

Table 1 Characteristics of nuclear and mitochondrial gene segments

Gene	Nucleotides			Seq. div. range (%)	Ident. mouse vs human (%)	CI-MP	No. of species amplified	Amino acids		
	No. of bp	No. of var. sites	No. of PI					No. of res.	No. of var. sites	No. of PI
ADORA3	330	227	191	0.31–37.08	78.5	0.34	63	110	80	62
ADRB2*	834	392	313	0.36–18.06	86.4	0.35	60	278	97	63
APP-3'UTR	690	454	299	0.32–31.16	90.5	0.51	63	—	—	—
ATP7A	690	468	375	0–41.66	87.1	0.41	64	230	167	129
BDNF	558	250	179	0.54–27.52	90.7	0.36	63	186	73	38
BM11-3'UTR	340	175	78	0–16.78	93.4	0.65	56	—	—	—
CNR1	1,002	398	332	0.41–27.25	89.6	0.30	62	334	55	24
CREM-3'UTR	422	302	227	0–28.53	85.1	0.46	65	—	—	—
EDG1	978	458	365	0.10–19.26	88.5	0.35	58	326	98	62
PLCB4-3'UTR	337	290	249	0.33–70.40	78.9	0.44	65	—	—	—
PNOC*	333	218	173	0–42.55	82.5	0.43	56	111	69	62
RAG1	774	353	303	0.78–23.98	87.1	0.32	56	258	66	47
RAG2*	444	255	203	0.45–23.36	89.2	0.38	62	148	86	54
TYR	426	258	224	0–34.66	85.9	0.36	52	142	83	67
ZFX	204	79	62	0–22.99	93.6	0.41	57	68	14	5
mtDNA	1,417	929	731	0–31.69	81.7	0.25	66	—	—	—
Total	9,779	5,506	4,304					2,416	963	655

* No outgroup included in comparisons.

bp, base pairs; var., variable sites; PI, phylogenetically informative sites; res., amino-acid residues; Seq. div., percentage of nucleotide sequence divergence (Kimura 2-parameter); CI, consistency index of most parsimonious tree(s).

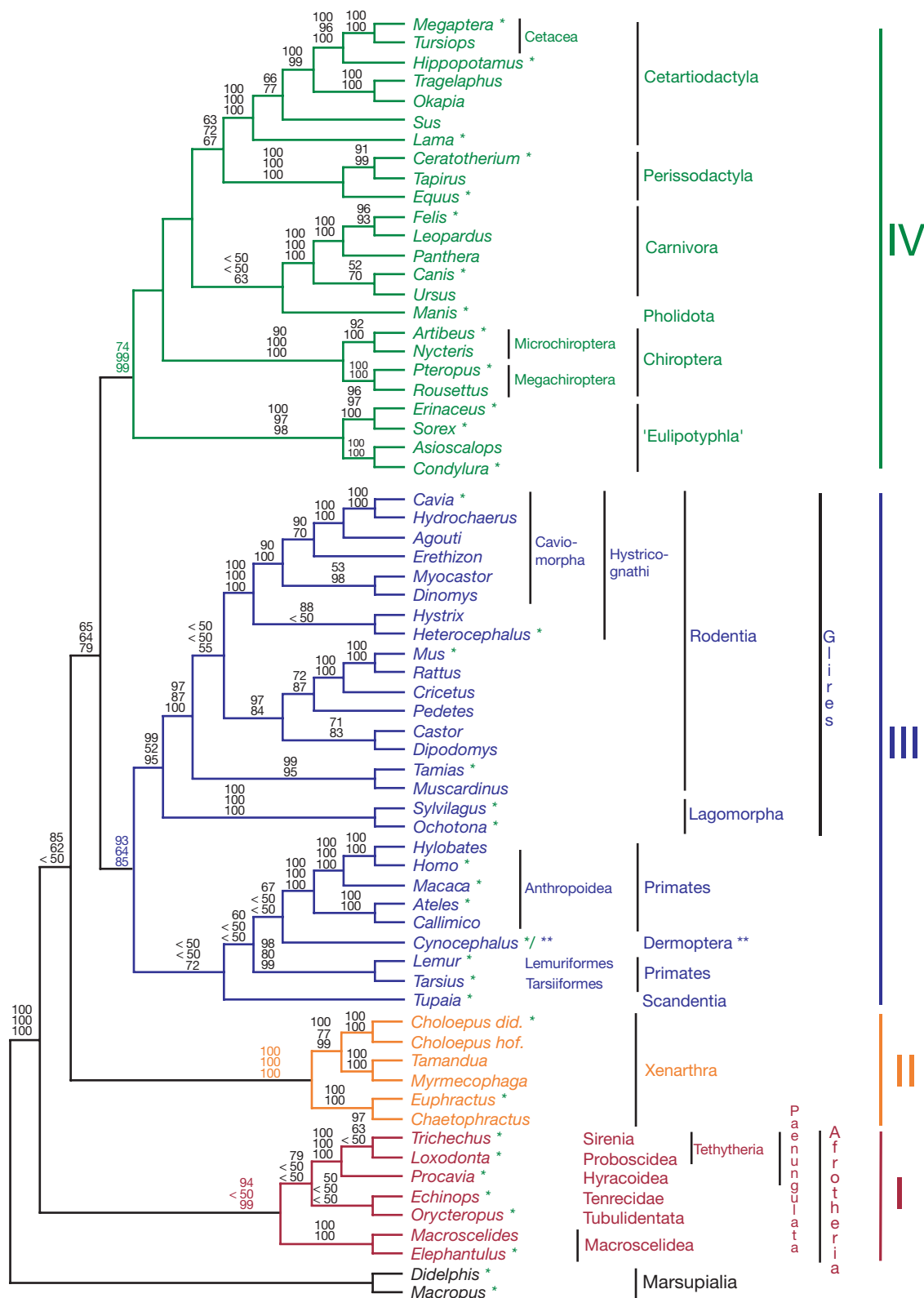


Figure 1 Phylogenetic relationships among 64 placental mammals and two marsupials based on analysis of 9,779 bp from 15 nuclear and three mtDNA genes. The tree represents the minimum evolution topology estimated through neighbour joining (NJ), using maximum likelihood (ML) distances (see Methods). Maximum parsimony (MP) and ML analyses (see Supplementary Information) produced a similar topology (MP: TL = 26,422, consistency index CI = 0.34, retention index RI = 0.47; ML: Ln Likelihood = -95086.03) with differences revolving mostly around short internodes with low bootstrap support. Numbers indicate per cent bootstrap support from NJ (top), MP (middle) and ML (bottom), based on 1,000 iterations for NJ and MP, and 100 iterations for ML. ML analyses were based on a pruned data set (37 taxa marked with asterisks). Terminal taxa

are labelled by their generic designation, except where multiple members of a genus were included (see Methods). Brackets indicate higher level taxonomic groups observed in resulting trees (right of tree). A difference between MP/ML and NJ analyses was the position of the flying lemur (*Cynocephalus*, double asterisk). MP and ML differed from the shown tree by supporting primate monophyly and a sister-group relationship between Dermoptera and Scandentia. Weighted parsimony analyses using only transversions (Tv), removal of third position transitions (Ts), and a Tv:Ts weight of 2:1 produced topologies congruent with the shown tree, with differences revolving only around branches depicted here with low (< 50%) bootstrap support.

relationship that has also been suggested by previous morphological and molecular data^{6,21,22}. Although our support for the Carnivora–Pholidota sister grouping fails statistical tests against alternative hypotheses (Kishino–Hasegawa (parsimony and likelihood) and Templeton tests; $P \geq 0.345$), our data do firmly place Pholidota within clade IV, and reject the view of an alliance between pangolins and the Neotropical Xenarthrans² (Kishino–Hasegawa (parsimony and likelihood) and Templeton tests, $P < 0.0017$). Our data also consistently indicated a sister-group relationship between Cetartiodactyla¹⁹ and Perissodactyla, albeit with only moderate bootstrap support (63–72%).

We also performed phylogenetic analyses with the combined amino-acid sequences from the 11 coding nuclear genes, which provided medium to high support for the monophyly of most orders, as well as some superordinal relationships (for example, Afrotheria, Paenungulata, Glires, and a split between Afrotheria plus Xenarthra versus other placentals; see Supplementary Information). However, these data lacked sufficient power to resolve the other basal relationships identified with the nucleotide data set. This is not unexpected, given that the amino-acid data set contains less than 20% of the number of variable and parsimony-informative sites present in the nucleotide data set (Table 1).

Notably, our data set shows consistency of amplification and phylogenetic signal associated with the four non-coding nuclear segments (*APP*, *BMI1*, *CREM* and *PLCB4*) all of which were derived from 3' untranslated regions (3' UTRs). Whereas the average percentage of variable sites was expectedly higher for the 3' UTRs (68% of 1,789 bp, $n = 4$) versus coding regions (51% of 6,573 bp, $n = 11$), the average consistency index was also higher for the 3' UTRs (0.51) compared with the coding regions (0.33). Within the UTRs, we generally observed conserved blocks of sequence interrupted by variable (yet usually alignable) indels, which were often phylogenetically informative. It is probable that 3' UTRs will be of importance in future phylogenetic studies of mammalian orders, as well as higher level phylogeny within other taxonomic groups.

The molecular resolution of placental mammals into four superordinal clades has been independently corroborated by an accompanying analysis²³. Estimates for the timing of molecular divergence indicate that the superordinal diversification occurred 64–104 Myr ago (median 84 Myr ago²⁴), several millions of years prior to the K/T boundary which marked the disappearance of the dinosaurs²⁵. The earliest divergences (Afrotheria and Xenarthra) apparently occurred in the Southern Hemisphere, probably associated with the geological separation of the southern supercontinent Gondwanaland, followed by dispersal of the ancestors of Clades III and IV into the northern supercontinent of Laurasia²⁴. The placement of several primitive insectivorous/generalist taxa near basal positions in each of the major clades (for example, Tenrecidae, Eulipotyphla and Scandentia; Fig. 1) is consistent with the early eutherian divergences preceding their remarkable morphological diversification. Geographical isolation of these primitive forms in different regions and the appearance of ecological niches vacated by the dinosaurs may account for the remarkable parallel adaptive radiations suggested by Madsen *et al.*²³

From a genomics standpoint, these findings suggest that mouse and human, the two species at the forefront of genomic sequencing, are phylogenetically restricted to just one of the four principal placental lineages identified in our analyses. A broader knowledge of mammalian genomic organization, function and evolution will be gained by applying large-scale genome analysis to other species from each of these principal lineages, particularly Afrotheria and Xenarthra. This will achieve a better understanding of the history of our mammalian ancestors and the uniqueness of our own genome. □

Methods

Taxon sampling, amplification and sequencing

Our data set contains representatives from divergent lineages of all eutherian orders. Order

Xenarthra: *Choloepus didactylus* (Linne's two-toed sloth), *Choloepus hoffmanni* (Hoffmann's two-toed sloth), *Tamandua tetradactyla* (tamandua), *Myrmecophaga tridactyla* (giant anteater), *Euphractus sexcinctus* (six-banded armadillo), *Chaetophractus villosus* (hairy armadillo); Order Insectivora: *Erinaceus concolor* (European hedgehog), *Sorex araneus* (European shrew), *Asioscalops altaica* (Old World mole), *Condylura cristata* (star-nose mole), *Echinops telfairi* (tenrec); Order Macroscelidea: *Elephantulus rufescens* (long-eared elephant shrew), *Macroscelides proboscideus* (short-eared elephant shrew); Order Tubulidentata: *Orycteropus afer* (aardvark); Order Hyracoidea: *Procavia capensis* (rock hyrax); Order Sirenia: *Trichechus manatus* (West Indian manatee); Order Proboscidea: *Loxodonta africana* (African elephant); Order Rodentia: *Tamias striatus* (eastern chipmunk), *Castor canadensis* (American beaver), *Muscardinus avellanarius* (dormouse), *Pedetes capensis* (springhare), *Mus musculus* (mouse), *Rattus norvegicus* (brown rat), *Cricetus griseus* (hamster), *Hystrix brachyurus* (Malayan porcupine), *Erethizon dorsatum* (North American porcupine), *Dipodomys heermanni* (kangaroo rat), *Heterocephalus glaber* (naked mole rat), *Cavia tschudii* (guinea-pig), *Hydrochaeris hydrochaeris* (capybara), *Myocastor coypus* (nutria), *Dinomys branickii* (pacarana), *Agouti taczanowskii* (mountain paca); Order Lagomorpha: *Ochotona hyperborea* (pika), *Sylvilagus cf. floridanus* (eastern cottontail); Order Dermoptera: *Cynocephalus variegatus* (Malayan flying lemur); Order Scandentia: *Tupaia minor* (lesser tree shrew); Order Primates: *Lemur catta* (ring-tailed lemur), *Tarsius bancanus* (Western tarsier), *Macaca mulatta* (rhesus macaque), *Ateles fusciceps* (brown-headed spider monkey), *Callimico goeldi* (Goeldi's monkey), *Hylobates concolor* (gibbon), *Homo sapiens* (human); Order Chiroptera: *Rousettus lanosus* (long-haired rousette), *Pteropus giganteus* (flying fox), *Nycteris thebaica* (slit-faced bat), *Artibeus jamaicensis* (Jamaican fruit-eating bat); Order Carnivora: *Canis familiaris* (domestic dog), *Ursus arctos* (brown bear), *Felis catus* (domestic cat), *Leopardus pardalis* (ocelot), *Panthera onca* (jaguar); Order Perissodactyla: *Ceratotherium simum* (white rhinoceros), *Equus caballus* (horse), *Tapirus indicus* (Malayan tapir); Order Artiodactyla: *Lama glama* (llama), *Sus scrofa* (pig), *Hippopotamus amphibius* (hippopotamus), *Okapia johnstoni* (okapi), *Tragelaphus euryceros* (bongo); Order Cetacea: *Megaptera novaeangliae* (humpback whale), *Tursiops truncatus* (bottlenose dolphin); Order Pholidota: *Manis pentadactyla* (Chinese pangolin). For outgroup comparison we included both Australian (*Macropus eugenii*, Tamar wallaby) and American (*Didelphis virginianus*, opossum) marsupial representatives.

We designed primers as part of an ongoing effort to develop markers with uses in both mammalian comparative gene mapping and mammalian phylogenetics. The GenBank and UniGene databases (NCBI) were searched for genes with exons of sufficient length (> 200 bp) and variability (80–95% nucleotide identity between mouse and human), thereby providing adequate variation for the purpose of phylogenetic and somatic cell/radiation hybrid mapping. We designed primers in conserved stretches of the genes using Primer 3.0 (Whitehead Institute). Primer design avoided regions showing BLAST similarity to paralogues of the target gene. Primer pairs were selected by pre-screening each set in a panel of DNA from 11 individuals, including ten species from different eutherian Orders and one marsupial, using a touchdown PCR protocol with *Taq*-Gold DNA polymerase (PE Biosystems), and either 1.5 or 2.0 mM MgCl₂. Only those markers that showed amplification of a single, intense product of the predicted size in at least nine of the ten eutherian orders were chosen for the final study. We then chose successful primer pairs (see Supplementary Information) for amplification in a 96-well format (2 × 48 species or 6 × 16 species). We amplified remaining species in separate reactions with appropriate controls. Amplification products and Big Dye terminator (Applied Biosystems) sequencing reactions were purified using the *Psiclon* 96-well PCR purification system (Princeton Separations), and the 96-well Centri-sep system (Princeton Separations), respectively. Sequencing reactions were largely performed in a 96-well format using the ABI-3700 capillary sequencing apparatus, whereas smaller numbers of reactions were analysed using either an ABI-373 stretch or ABI-377 DNA sequencer.

Our initial data set consisted of > 12,000 bp from 20 nuclear-coding loci. We removed five genes (1,672 bp) owing to inconsistent amplification in > 20% of the species, failure of key taxa, or amplification of putative paralogous loci in some species. The final set of genes comprised 10,704 bp of which 925 bp were deleted because of ambiguous alignment, resulting in the total data set of 9,779 bp. A summary of the genes used and their characteristics is given in Table 1.

Data alignment and phylogenetic analysis

Phylogenetic analysis was performed on data sets aligned using the computer software CLUSTAL-X²⁶. We manually inspected and modified alignment output. Regions of the alignments in which determination of homology was ambiguous were deleted before phylogenetic analyses. In the final data set, all taxa were represented by at least 11 of the 17 gene segments (see Supplementary Information). Phylogenetic analyses were performed using maximum parsimony, distance-based (neighbour joining), and maximum likelihood methods implemented in PAUP* 4.0b2 (ref. 27), PHYLIP 3.57, (DNAML and Protdist, J. Felsenstein, Univ. of Washington) and PUZZLE²⁸. PAML 2.0 (Z. Yang, University College, London) was used for estimating the transition/transversion ratio ($Ts/Tv = 2.0$) and the α -parameter for rate variation among sites ($\alpha = 0.45$). Parsimony analyses were performed using 50 replicates with random addition of taxa and tree-bisection reconnection branch swapping.

We performed minimum evolution (neighbour joining) analyses with several distance measures (Kimura 2-parameter, Logdet and maximum likelihood distances with HKY85 model) to examine effects on topological stability (see Supplementary Information). Maximum likelihood analyses for the nucleotide data set were performed with DNAML (assuming equal substitution rates among sites) and with PAUP* (using an HKY- γ model with estimated parameters); maximum likelihood trees for the amino-acid data set were constructed using PUZZLE (γ -corrected JTT model). Reliability of nodes was assessed using 1,000 bootstrap iterations for the nucleotide maximum parsimony and neighbour-

joining trees and the amino-acid maximum parsimony phylogenies, and 100 replicates for the nucleotide maximum likelihood tree and the amino-acid distance-based analyses (Dayhoff PAM matrix) (see Supplementary Information for additional trees and summary of bootstrap support). We performed tests of alternative phylogenetic hypotheses using Kishino–Hasegawa²⁹ (parsimony and likelihood) and Templeton's non-parametric³⁰ tests.

Received 30 October; accepted 4 December 2000.

- Eisenberg, J. F. *The Mammalian Radiations* (Chicago Univ. Press, Chicago, 1981).
- Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356**, 121–125 (1992).
- O'Brien, S. J. *et al.* The promise of comparative genomics in mammals. *Science* **286**, 458–481 (1999).
- Springer, M. S. *et al.* Endemic African mammals shake the phylogenetic tree. *Nature* **388**, 61–64 (1997).
- Stanhope, M. J. *et al.* Highly congruent molecular support for a diverse clade of endemic African mammals. *Mol. Phylogenet. Evol.* **9**, 501–508 (1998).
- McKenna, M. C. & Bell, S. K. *Classification of Mammals above the Species Level* (Columbia Univ. Press, New York, 1997).
- Mouchatty, S. K., Gullberg, A., Janke, A. & Arnason, U. The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. *Mol. Biol. Evol.* **17**, 60–67 (2000).
- Stanhope, M. J. *et al.* Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl Acad. Sci. USA* **95**, 9967–9972 (1998).
- Simpson, G. G. The principles of classification and a classification of mammals. *Bull. Am. Mus. Nat. Hist.* **85**, 1–350 (1945).
- McKenna, M. C. in *Phylogeny of the Primates: a Multidisciplinary Approach* (eds Luckett, W. P. & Szalay, F. S.) 21–46 (Plenum, New York, 1975).
- Easteal, S. The pattern of mammalian evolution and the relative rate of molecular evolution. *Genetics* **124**, 165–173 (1990).
- Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
- Reyes, A. *et al.* Where do rodents fit? Evidence from the complete mitochondrial genome of *Sciurus vulgaris*. *Mol. Biol. Evol.* **17**, 979–983 (2000).
- Gregory, W. K. The orders of mammals. *Bull. Am. Mus. Nat. Hist.* **27**, 1–524 (1910).
- Graur, D., Duret, L. & Gouy, M. Phylogenetic position of the order Lagomorpha (rabbits, hares and allies). *Nature* **379**, 333–335 (1996).
- Huchon, D., Catzeflis, F. M. & Douzery, E. J. P. Molecular evolution of the nuclear von Willebrand factor gene in mammals and the phylogeny of rodents. *Mol. Biol. Evol.* **16**, 577–589 (1999).
- Kay, R. F., Ross, C. & Williams, B. A. Anthropoid origins. *Science* **275**, 797–804 (1997).
- Pettigrew, J. D. Flying Primates? Megabats have advanced pathway from eye to midbrain. *Science* **231**, 1304–1306 (1986).
- Waddell, P. J., Okada, N. & Hasegawa, M. Towards resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48**, 1–5 (1999).
- Krettek, A., Gullberg, A. & Arnason, U. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Eriacus europaeus*, and the phylogenetic position of the Lipotyphla. *J. Mol. Evol.* **41**, 952–957 (1995).
- Rose, K. D. & Emry, R. J. in *Mammal Phylogeny: Placentals* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 81–102 (Springer, New York, 1993).
- Miyamoto, M. M. & Goodman, M. Biomolecular systematics of eutherian mammals: phylogenetic patterns and classification. *Syst. Zool.* **35**, 230–240 (1986).
- Madsen, O. *et al.* Parallel adaptive radiations in two major clades of placental mammals. *Nature* (in the press).
- Eizirik, E., Murphy, W. J. & O'Brien, S. J. Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* (in the press).
- Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229 (1996).
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* **25**, 4876–4882 (1997).
- Swofford, D. L. *PAUP* Phylogenetic Analysis using Parsimony and Other Methods* (Sinauer, Sunderland, Massachusetts, 1998).
- Strimmer, K. & Von Haeseler, A. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**, 964–969 (1996).
- Kishino, H. & Hasegawa, M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**, 170–179 (1989).
- Templeton, A. R. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37**, 221–244 (1983).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature. Sequences are deposited in GenBank under accession numbers AY011125–AY012154.

Acknowledgements

We thank D. Hirschmann, M. Houck, R. Montali, R. Baker, G. Harris, K. Helgen, A. L. Roca, M. Roelke-Parker, A. Grafodatsky, O. Serov and T. Oleksyk for help in obtaining samples and technical assistance, and M. Smith and M. Dean for helpful suggestions. We also thank the NCI Frederick Molecular Technology Center for technical support, and the Advanced Biomedical Computer Center for computational assistance. All tissue samples were obtained with appropriate permits (CITES) issued to the National Cancer Institute, National Institutes of Health (principal officer, S.J.O.). Y.P.Z. is supported by the Natural Science Foundation of China and Chinese Academy of Sciences. E.E. is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

Correspondence and requests for materials should be addressed to S.J.O. (e-mail: obrien@mail.ncifcrf.gov).

Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants

Kathleen M. Pryer*, Harald Schneider*, Alan R. Smith†, Raymond Cranfill†, Paul G. Wolf‡, Jeffrey S. Hunt* & Sedonia D. Sipes‡

* Department of Botany, The Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, Illinois 60605, USA

† University Herbarium, University of California, 1001 Valley Life Sciences Building 2465, Berkeley, California 94720, USA

‡ Department of Biology, 5305 Old Main Hill, Utah State University, Logan, Utah 84322, USA

Most of the 470-million-year history of plants on land belongs to bryophytes, pteridophytes and gymnosperms, which eventually yielded to the ecological dominance by angiosperms 90 Myr ago^{1–3}. Our knowledge of angiosperm phylogeny, particularly the branching order of the earliest lineages, has recently been increased by the concurrence of multigene sequence analyses^{4–6}. However, reconstructing relationships for all the main lineages of vascular plants that diverged since the Devonian period has remained a challenge. Here we report phylogenetic analyses of combined data—from morphology and from four genes—for 35 representatives from all the main lineages of land plants. We show that there are three monophyletic groups of extant vascular plants: (1) lycophytes, (2) seed plants and (3) a clade including equisetophytes (horsetails), psilotophytes (whisk ferns) and all eusporangiate and leptosporangiate ferns. Our maximum-likelihood analysis shows unambiguously that horsetails and ferns together are the closest relatives to seed plants. This refutes the prevailing view that horsetails and ferns are transitional evolutionary grades between bryophytes and seed plants⁷, and has important implications for our understanding of the development and evolution of plants⁸.

Estimates of a phylogeny for the main groups of land plants, each with highly divergent morphologies, have been many, and all have been contested. Bryophytes (liverworts, hornworts and mosses) are consistently shown to be a basal grade, but their relationships to one another and to vascular plants are controversial^{1,2,9–13}. Most phylogenetic analyses of vascular plants consistently reconstruct two main lines of evolution: the Lycopytina (clubmosses and relatives), with 1% of extant diversity, and the Euphyllophytina (all other vascular plants)^{1,2,10,11,14–17}. Extant Euphyllophytina^{1,2} comprises six major monophyletic lineages: Equisetopsida (horsetails), Polypodiidae (leptosporangiate ferns), Spermatophytata (seed plants), Psilotidae (whisk ferns; simple plants regarded by some to be living relicts of the earliest vascular plants^{7,18}), Marattiidae and Ophioglossidae (eusporangiate ferns). Phylogenetic assessments based on single genes^{10,11,14,15,19} and/or morphology^{1,7,12,17,20} have provided only weak and usually contradictory evidence for the relationships among these euphyllophyte lineages. Resolving these relationships would not only stabilize a pivotal region of vascular plant phylogeny but is also key to identifying the most appropriate outgroup for addressing questions related to the evolution of seed plants.

Recent palaeontological studies^{1,2,7} attempted to demonstrate that approaches based solely on living species would have difficulties reconstructing relationships among major lineages of vascular plants. Inadequate taxon sampling, rate heterogeneity across DNA nucleotide sites among lineages, and inappropriate algorithms also have been cited as impediments to resolving ancient branching events²¹. However, as predicted by recent theoretical studies²², combined analysis of DNA sequences from multiple loci proves to

be very useful in inferring deep phylogenetic patterns^{4–6}. With few exceptions^{12,20}, broad phylogenetic studies rely solely on combined nucleotide sequence data, with authors arguing that morphological character homology assessment among ancient and divergent groups is too challenging. This practice ignores the higher complexity of morphological characters that can conserve character states over time and that have a lower probability of random evolution of similar structures.

We obtained DNA sequences (5,072 aligned base pairs) of four genes from two plant genomes: plastid *atpB*, *rbcL* and *rps4*, and nuclear small-subunit ribosomal DNA. We also assembled a congruent data set of 136 vegetative and reproductive morphological/anatomical characters. We sampled 35 representatives from all major monophyletic lineages of land plants. The selection of taxa reflects our focus on basal vascular plants, and all six Euphyllophytina¹ lineages are represented by two or more members. Five bryophytes

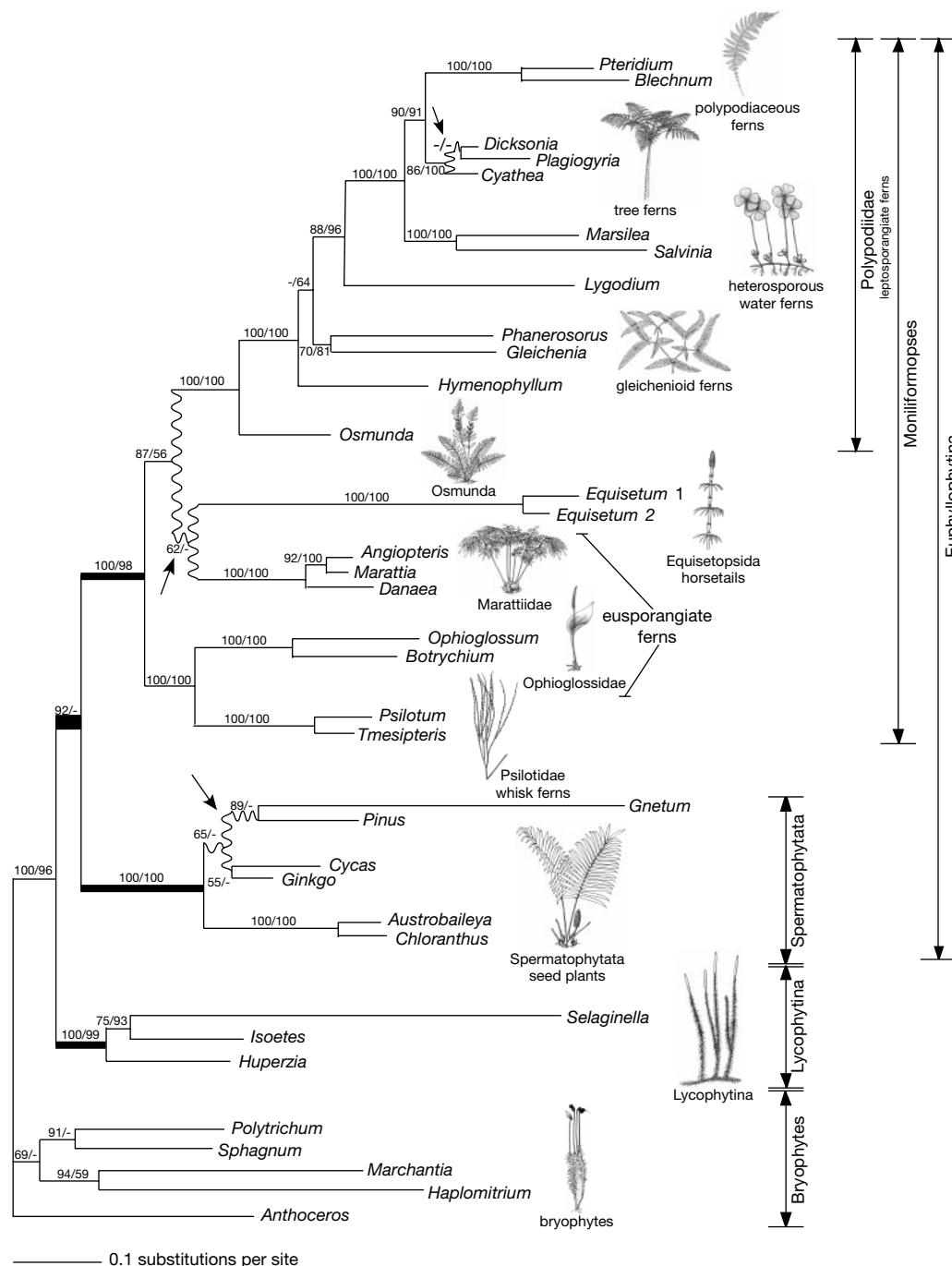


Figure 1 Phylogenetic relationships for all the main lineages of vascular plants inferred from maximum-likelihood (ML) analysis of the combined chloroplast *rbcL*, *atpB*, *rps4* and nuclear small-subunit rDNA data set. Numbers at nodes and before the slash are ML bootstrap values $\geq 50\%$; maximum parsimony (MP) bootstrap values $\geq 50\%$ appear after the slash when these same nodes were supported in the MP unequally weighted analysis of the combined four-genes plus morphology data set (single MP tree = 14165.04 steps). A minus sign indicates a node had less than 50% bootstrap support in one or the other analysis. The topology is rooted by all bryophytes, hence relationships depicted among

them are arbitrary. Branches leading to the three monophyletic clades of vascular plants (lycophytes, seed plants and horsetails+ferns) are drawn medium thick. The branch supporting the Euphyllophytina, with horsetails+ferns as sister group to seed plants, is the thickest. Wiggled lines (at straight arrows) indicate three areas of conflict between the ML and MP analyses. Branch lengths are proportional to number of substitutions per site (scale bar). Thumbnail sketches of plant representatives accompany major clades. Taxonomy follows ref. 1.

were specified as outgroups. We analysed the data sets using both maximum-parsimony (MP) and maximum-likelihood (ML) optimization criteria; bootstrap (BS) analyses were conducted to measure the stability of observed phylogenetic patterns.

Using ML on the combined four-gene data set we recovered one most likely tree ($-\ln$ likelihood = 36466.6365) for each of the 100 replicates (Fig. 1). We also observed an essentially identical topology using MP on the combined four-gene and morphology data set (three areas that differ are highlighted on Fig. 1). Regardless of the analytical approach (MP or ML), three major lineages emerged as monophyletic clades with exceptional support (100% BS). The first clade comprises the Lycophytina, increasingly recognized as a distinct group of vascular plants only distantly related to other extant pteridophytes and seed plants^{1,16}. The second diverging lineage corresponds to seed plants. The third, novel, clade includes all non-seed-producing lineages of Euphyllophytina, including horsetails (*Equisetopsida*), leptosporangiate ferns (*Polypodiidae*), eusporangiate ferns (*Marattiidae*, *Ophioglossidae*) and whisk ferns (*Psilotidae*). Seed plants, ferns and horsetails are united as a monophyletic group, to the exclusion of lycopods, in both the ML (92% BS) and MP (<50% BS) analyses.

We observed one unambiguous length discrepancy in *rps4* that can be interpreted as a molecular 'signature' providing additional support for horsetail-fern monophyly. A portion of the *rps4* alignment is shown for base pairs 646–696 (Fig. 2), which includes 27 ambiguously aligned base pairs (658–684) flanked by unambiguously aligned regions. The ambiguously aligned region was excluded entirely from the ML analysis. In the MP analysis, the same region was recoded simply as a single absence/presence character for the observed length increase. This multi-residue length increase in horsetails and ferns is not as likely to be a random convergence as is a single point mutation and provides further evidence for this clade.

Within the horsetail-fern lineage, *Psilotidae* is most closely related to *Ophioglossidae* (100% BS). Although this association was only weakly suggested in recent single-gene analyses^{11,19,20}, the current evidence unambiguously invalidates the traditional morphological and palaeobotanical view that *Psilotidae* are relatively unaltered descendants of the psilotophytes, among the earliest vascular plant fossils^{7,18}. *Ophioglossidae* and *Psilotidae* differ so radically in phenotype that this close relationship, implying a shared origin of phenotypic simplification, was never before explicitly considered. All other ferns and horsetails make up its sister clade (87% BS). The relationships of horsetails also have been controversial: sister to seed plants⁷, sister to leptosporangiate (*Polypodiidae*) and eusporangiate (*Ophioglossidae* and *Marattiidae*) ferns¹, or as a basal grade euphyllophyte lineage¹⁷. Our analysis clearly (100% BS) places *Equisetum* within the non-lycophyte pteridophyte clade, although its exact relationships within this clade are not yet well resolved. In the ML analysis, *Equisetum* is sister to *Marattiidae* (62% BS), whereas in the MP analysis, it is sister to leptosporangiate ferns (<50% BS). This study also confirms a sister relationship between tree ferns and the more derived 'polypodiaceous' leptosporangiate ferns (90% BS), and places the heterosporous water ferns as sister to this clade (100% BS) (Fig. 1). Relationships among these groups were equivocal in earlier studies^{17,20}.

The only noteworthy disagreement between our ML and MP analyses is localized within seed-plant relationships, a subject of much current controversy^{21,23,24}. Our ML analysis resolved gymnosperms as monophyletic (65% BS) and *Gnetum* as sister to *Pinus* (89% BS). Our MP analysis supports *Gnetum* as basal among seed plants (87% BS), and all other gymnosperms as monophyletic (67% BS) and sister to angiosperms.

In the ML analysis of the combined four-gene data set, there is persuasive support for the Euphyllophytina (92% BS), with a basal dichotomy indicating that the horsetail-fern clade (100% BS) is the closest relative to seed plants (100% BS). To the best of

our knowledge, this relationship has been proposed only once previously¹, as a tentative hypothesis on the evidence of a single anatomical character (protoxylem distribution). This led to the provisional classification of the horsetail-fern clade as infradivision Moniliformopses (moniliforms); *Psilotidae*, however, was not included in that study¹. Although this same deep dichotomy is also robustly resolved in the MP analysis of the combined four-genes plus morphology data set, the Euphyllophytina node is weakly supported (<50% BS). Exceptionally long branches in each of the three main clades (Fig. 1: *Selaginella*, *Gnetum* and *Equisetum*) and the greater sensitivity of MP over ML to long-branch attraction (statistical inconsistency) effects^{21,25} probably explain why parsimony bootstrapping failed to recover this clade with high confidence. When these long-branch taxa were removed and the combined four-genes plus morphology data set was re-analysed with MP, this same basal Euphyllophytina node was highly supported (83% BS, results not shown). Each of our separate single-gene analyses, with the exception of *rps4*, did not resolve the horsetail-fern clade, and none was able to determine confidently the closest relatives to seed plants. Only our morphological data set, when analysed alone with MP, provided the same conclusions

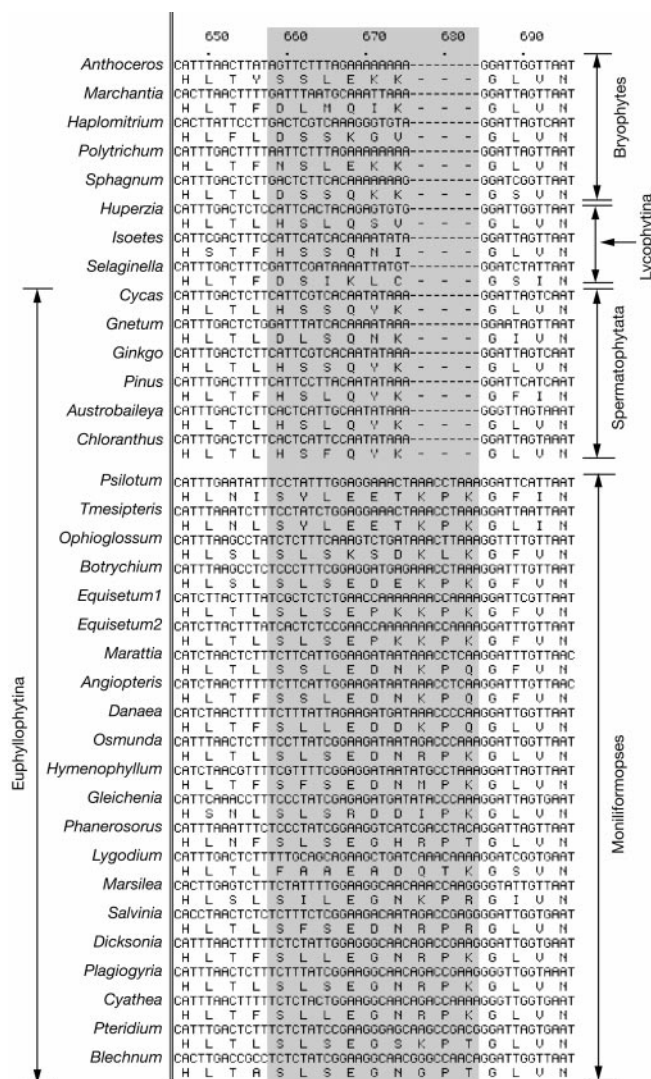


Figure 2 A portion of the chloroplast *rps4* alignment. An ambiguously aligned region (grey box) containing a 9-base-pair length difference distinguishes horsetails and ferns (bottom block) from bryophytes, lycophytes and seed plants (top block). Amino-acid translations are interleaved below each DNA sequence. Dashes indicate gaps.

regarding the Euphyllophytina as when the four genes were analysed simultaneously with ML. A study using mitochondrial small-subunit rDNA sequence data¹⁰ with a smaller selection of taxa suggested support for this hypothesis; however, critical euphyllophyte taxa (Psilotidae and Marattiidae) were not included. A more recent study²⁶ that combines data from two genes (nuclear and mitochondrial small-subunit rDNA) strongly corroborates a horsetail–fern clade as sister to seed plants, despite a limited sampling of only seven euphyllophyte taxa from all pertinent lineages.

Our report of a basal dichotomy in the Euphyllophytina, a split that occurred in the early–mid Devonian (about 400 Myr ago), necessitates abandonment of the prevailing view that ferns and horsetails represent paraphyletic successive grades of increasing complexity in early vascular plant evolution, which eventually led to the more complex seed plants, and ultimately to angiosperms. A parallel deep reorganization of metazoan phylogeny has recently been proposed²⁷, with ‘simple’ bilaterian taxa (for example, platyhelminths and nemerteans) being displaced from the base of the metazoan tree to within the large lophotrochozoan clade.

A corollary of the demise of the paraphyletic interpretation of early vascular plant evolution is that it is now necessary to confront the many recurring models that derive morphological, developmental and physiological conditions in seed plants from an ‘intermediate’ or ‘primitive’ pteridophyte ancestor. We predict that this will require a significant revision in the interpretation of the underlying processes of vascular plant evolution. For example, a number of homeotic genes, such as the MADS-box genes that encode transcription factors critical for regulating physiological and developmental processes, especially flower development, have been well studied in angiosperms²⁸. Clarifying the origin of these genes has been hampered by the few reports of homologues from non-seed plants, and therefore it is not known to what extent changes in number, regulation and function of these and other homeotic genes may have driven land plant evolution. The study of these genes from across a stable phylogenetic framework is critical. We note that all the main plant model organisms (for example, *Arabidopsis*, *Glycine*, *Lycopersicon*, *Oryza*, *Petunia* and *Zea*) are recently evolved angiosperms. Efforts to promote developmental and genomic research on model systems in the horsetail–fern clade (for example, *Ceratopteris*²⁹), will probably lead to an improved understanding of fundamental aspects of vascular plant development and evolution⁸. □

Methods

Taxon sampling and morphological data set

We selected 35 taxa to sample explicitly at least two members of each major monophyletic group of land plants. The various groups were determined from recent broad-scale phylogenetic analyses^{1,12,17,20}, and we specified the bryophytes *Anthoceros*, *Haplomitrium*, *Marchantia*, *Polytrichum* and *Sphagnum* as outgroups. Our morphological data set comprised 136 parsimony-informative characters (H.S. *et al.*, manuscript in preparation), which we, for the most part, adopted or modified from recent studies^{1,12,17,20}.

Gene sequencing

We amplified chloroplast *rbcl*, *atpB*, *rps4*, and nuclear small-subunit rDNA genes for all 35 taxa from total cellular DNA by polymerase chain reaction (PCR) and sequenced them using an ABI 377 automated DNA sequencer (PE Applied Biosystems). Details of taxon sampling, DNA isolation, PCR amplification, sequencing, sequence alignment, exclusion and recoding of ambiguously aligned regions, data set combinability testing, and phylogenetic analysis will be published elsewhere (K.M.P. *et al.*, manuscript in preparation). Most *atpB*, *rps4*, nuclear small-subunit rDNA, and some *rbcl* sequences were generated as part of this study. For voucher information, GenBank numbers and the aligned data matrices, see Supplementary Information and http://www.fnmh.org/research_collections/botany/botany_sites/ferns/publications.html; data matrices are also available in TreeBASE, accession number S543, at <http://www.herbaria.harvard.edu/treebase/>.

Phylogenetic analyses

We conducted heuristic MP (unequal weighting schemes, 1,000 random-addition replicates, tree bisection–reconnection (TBR) branch swapping) and ML (general time-reversible model, accommodating unequal nucleotide frequencies and different

probabilities for each of six substitution types, plus three heterogeneous rate categories across sites following a discrete approximation of the gamma distribution, 100 random-addition replicates) analyses using PAUP* version 4.0b2a³⁰. The ML analysis was restricted to the combined four-gene data set because it is not possible to simultaneously implement two models of evolution, one for morphology and one for DNA sequence data, in any currently available computer programs. We further performed both parsimony bootstrap (unequal weighting schemes, 1,000 replicates, each with 10 random-addition replicates and TBR branch swapping) and likelihood bootstrap analyses (212 replicates, using identical parameters to those used to find the most likely tree).

Received 25 July; accepted 27 November 2000.

- Kenrick, P. & Crane, P. R. *The Origin and Early Diversification of Land Plants: A Cladistic Study* (Smithsonian Institution Press, Washington DC, 1997).
- Kenrick, P. & Crane, P. R. The origin and early evolution of plants on land. *Nature* **389**, 33–39 (1997).
- Lupia, R., Lidgard, S. & Crane, P. R. Comparing palynological abundance and diversity: Implications for biotic replacement during the Cretaceous angiosperm radiation. *Paleobiology* **25**, 305–340 (1999).
- Soltis, P. S., Soltis, D. E. & Chase, M. W. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**, 402–404 (1999).
- Qiu, Y.-L. *et al.* The earliest angiosperms: Evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404–407 (1999).
- Barkman, T. J. *et al.* Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc. Natl Acad. Sci. USA* **97**, 13166–13171 (2000).
- Rothwell, G. W. Fossils and ferns in the resolution of land plant phylogeny. *Bot. Rev.* **65**, 188–218 (1999).
- Graham, L. E., Cook, M. E. & Busse, J. S. The origin of plants: Body plan changes contributing to a major evolutionary radiation. *Proc. Natl Acad. Sci. USA* **97**, 4535–4540 (2000).
- Bremer, K. Summary of green plant phylogeny and classification. *Cladistics* **1**, 369–385 (1985).
- Duff, R. J. & Nickrent, D. L. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *Am. J. Bot.* **86**, 372–386 (1999).
- Hedderon, T. A., Chapman, R. & Cox, C. J. in *Bryology for the Twenty-first Century* (eds Bates, J. W., Ashton, N. W. & Duckett, J. G.) 65–77 (Maney Publishing and the British Bryological Society, Leeds, 1998).
- Mishler, B. D. *et al.* Phylogenetic relationships of the “green algae” and “bryophytes”. *Ann. Missouri Bot. Gard.* **81**, 451–483 (1994).
- Qiu, Y.-L., Cho, Y., Cox, J. C. & Palmer, J. D. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* **394**, 671–674 (1998).
- Hiesel, R., von Haeseler, A. & Brennicke, A. Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proc. Natl Acad. Sci. USA* **91**, 634–638 (1994).
- Kranz, H. D. & Huss, V. A. R. Molecular evolution of pteridophytes and their relationships to seed plants: Evidence from complete 18S rRNA gene sequences. *Plant Syst. Evol.* **202**, 1–11 (1996).
- Raubeson, L. A. & Jansen, R. K. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* **255**, 1697–1699 (1992).
- Stevenson, D. W. & Loconte, H. in *Pteridology in Perspective* (eds Camus, J. M., Gibby, M. & Johns, R. J.) 435–467 (Royal Botanic Gardens, Kew, 1996).
- Wagner, W. H. Jr Systematic implications of the Psilotaceae. *Brittonia* **29**, 54–63 (1977).
- Manhart, J. R. Phylogenetic analysis of green plant *rbcl* sequences. *Mol. Phylogenet. Evol.* **3**, 114–127 (1994).
- Pryer, K. M., Smith, A. R. & Skog, J. E. Phylogenetic relationships of extant ferns based on evidence from morphology and *rbcl* sequences. *Am. Fern J.* **85**, 205–282 (1995).
- Sanderson, M. J., Wojciechowski, M. F., Hu, J. M., Sher Khan, T. & Brady, S. G. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Mol. Biol. Evol.* **17**, 782–797 (2000).
- Hillis, D. M. Inferring complex phylogenies. *Nature* **383**, 130–131 (1996).
- Bowe, L. M., Coat, G. & dePamphilis, C. W. Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales’ closest relatives are conifers. *Proc. Natl Acad. Sci. USA* **97**, 4092–4097 (2000).
- Chaw, S.-M., Parkinson, C. L., Cheng, Y., Vincent, T. M. & Palmer, J. D. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proc. Natl Acad. Sci. USA* **97**, 4086–4091 (2000).
- Lewis, P. O. in *Molecular Systematics of Plants II: DNA Sequencing* (eds Soltis, D. E., Soltis, P. S. & Doyle, J. J.) 132–163 (Kluwer Academic, Boston, 1998).
- Renzaglia, K. S., Duff, R. J., Nickrent, D. L. & Garbary, D. J. Vegetative and reproductive innovations of early land plants: Implications for a unified phylogeny. *Phil. Trans. R. Soc. Lond. B* **355**, 769–793 (2000).
- Adoutte, A. *et al.* The new animal phylogeny: Reliability and implications. *Proc. Natl Acad. Sci. USA* **97**, 4453–4456 (2000).
- Theissen, G. *et al.* A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42**, 115–149 (2000).
- Hickok, L. G., Warne, T. R. & Fribourg, R. S. The biology of the fern *Ceratopteris* and its use as a model system. *Int. J. Plant Sci.* **156**, 332–345 (1995).
- Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other Methods)* (Sinauer, Sunderland, Massachusetts, 1999).

Supplementary information is available on Nature’s World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. Lupia, F. M. Lutzoni, B. D. Mishler, L. Newstrom-Lloyd and S. Zoller for critical comments on the manuscript; Z. Dabich, J. Bélisle, R. Lupia and D. Kieffer for assistance in rendering Fig. 1; F. M. Lutzoni and V. A. Funk for advice on phylogenetic analyses; I. Capesius, S. Boyles, B. Goffinet, M. Hasebe, M. Kato, M. Kessler, B. D. Mishler, R. Moran, J. Shaw, W. C. Taylor, Y.-L. Qiu, D. Wall, J. Wheeler, and greenhouse managers at Humboldt State University, University of California at Davis, University of California Botanical Garden at Berkeley, New York Botanical Garden for plant material; S. W.

Graham, P. S. Soltis and J. Therrien for sharing unpublished sequence data; and D. Ferguson, E. Grismer, J. Irwin and L. Sappelsa for general assistance in the initial stages of the project. This work was supported by grants from the NSF to K.M.P., A.R.S., P.G.W. and R. C., the Green Plant Phylogeny Research Coordination Group (USDA grant), and by the Pritzker Foundation Fund of The Field Museum.

Correspondence and requests for materials should be addressed to K.M.P. (e-mail: kpriyer@fieldmuseum.org).

Export by red blood cells of nitric oxide bioactivity

John R. Pawloski, Douglas T. Hess & Jonathan S. Stamler

Howard Hughes Medical Institute and Department of Medicine, Box 2612, Duke University Medical Center, Durham, North Carolina 27710, USA

Previous studies support a model in which the physiological O_2 gradient is transduced by haemoglobin into the coordinate release from red blood cells of O_2 and nitric oxide (NO)-derived vasoactivity to optimize oxygen delivery in the arterial periphery^{1,2}. But whereas both O_2 and NO diffuse into red blood cells, only O_2 can diffuse out^{3–5}. Thus, for the dilation of blood vessels by red blood cells, there must be a mechanism to export NO-related vasoactivity, and current models of NO-mediated intercellular communication should be revised. Here we show that in human erythrocytes haemoglobin-derived S-nitrosothiol (SNO), generated from imported NO, is associated predominantly with the red blood cell membrane, and principally with cysteine residues in the haemoglobin-binding cytoplasmic domain of the anion exchanger AE1. Interaction with AE1 promotes the deoxygenated structure in SNO-haemoglobin, which subserves NO group transfer to the membrane. Furthermore, we show that vasodilatory activity is released from this membrane precinct by deoxygenation. Thus, the oxygen-regulated cellular mechanism that couples the synthesis and export of haemoglobin-derived NO bioactivity operates, at least in part, through formation of AE1-SNO at the membrane-cytosol interface.

As the first step in analysing the fate of haemoglobin (Hb)-derived NO *in situ*, we determined the disposition of NO groups transferred physiologically from the haems of Hb to β -chain Cys 93 in intact human erythrocytes^{3,4}. Red blood cells (RBCs) held at less than 1% O_2 were exposed for 5 min to physiological amounts of NO (100 nM to 1 μ M; NO:haem ratios 1:1,000 to 1:100) followed by reoxygenation (21% O_2), and membrane and cytosolic fractions were prepared. Fractions were solubilized with Triton X-100 (TX100), and the NO content of extracts was measured by photolysis/chemiluminescence^{3,4}. At the lower NO:haem ratios, which produced intracellular NO concentrations matching those found *in vivo* (100–800 nM), recovery of NO was essentially complete, that is, none was lost to nitrate (Fig. 1a). In this model system, about 15–20% of NO incorporated by RBCs was present as SNO; the remainder was ascribed largely to iron nitrosyl haem (FeNO)^{1,3,4,6}. Most iron nitrosyl Hb was recovered with the cytosolic fraction (Fig. 1b). In contrast, SNO was associated predominantly with the membrane fraction (Fig. 1c). These results confirm that, in intact RBCs⁷ as with isolated reactants^{3,4}, Hb will efficiently capture and preserve NO, and form SNO, under physiological conditions. Unexpectedly, however, the formation of SNO is compartmentalized within the RBC.

Haemoglobin associates with the cytoplasmic face of the RBC membrane through specific protein-protein interactions^{8–10}. To determine the disposition of Hb-derived membrane SNO, we

examined the interaction of SNO-Hb^{5,6} with inside-out vesicles (IOVs) prepared by everting RBC membrane ghosts¹¹. IOVs incubated with SNO-Hb and washed at pH 8 to remove bound Hb incorporated about 450 pmol NO per mg of TX100-extracted IOV protein (Fig. 1d). All the incorporated NO was present in complex with thiol, that is, as SNO. It is important to note that SNO was not detected in extracts of IOVs exposed to NO in the absence of Hb (data not shown).

To rule out the possibility that apparent NO group transfer to IOVs was an artefact of residual membrane-bound SNO-Hb, we incubated IOVs with SNO-Hb immobilized on Sephadex beads. After centrifugal separation, washes at pH 7 and solubilization in TX100, extracts of IOVs were free of Hb as assessed by spectrophotometric detection of haem. SNO was present in those extracts at somewhat higher levels than in extracts derived from IOVs incubated with free SNO-Hb (suggesting a greater loss of

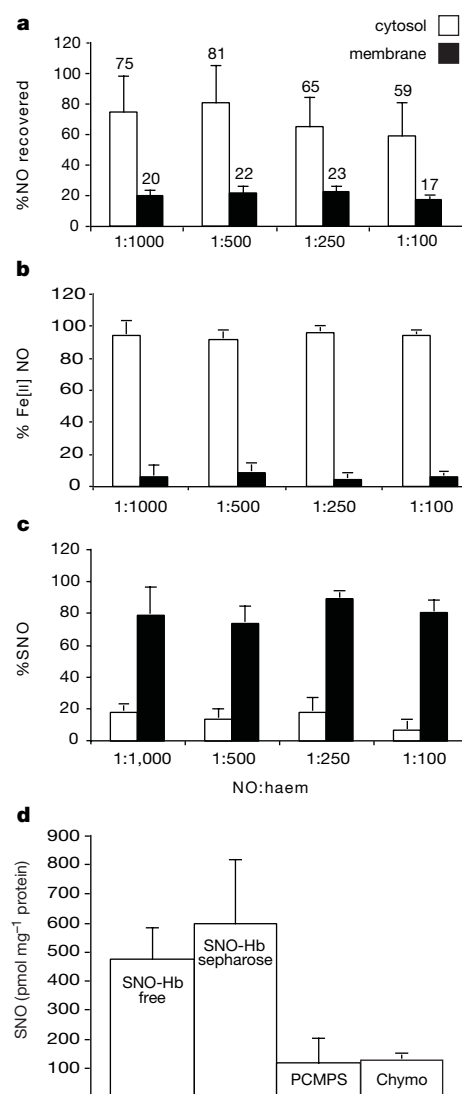


Figure 1 Haemoglobin-derived SNO is associated with cysteine thiols of RBC membrane proteins. **a–c**, Distribution in cytosolic and membrane fractions of NO groups after exposure of intact RBCs to NO. Recovery of NO is essentially complete at low, physiological NO:haem ratios (**a**), which yield 100–800 nM intracellular NO; FeNO is predominantly cytosolic (**b**), whereas SNO is largely membrane associated (**c**) ($P < 0.05$ for all pairwise comparisons). **d**, SNO content of IOVs exposed to free or Sephadex-bound SNO-Hb (50 nmol SNO-Hb per mg IOV protein). Transfer of NO groups to the membrane is greatly reduced ($P < 0.05$) after treatment of IOVs with the thiol-modifying reagent PCMPs and after mild digestion of IOVs with chymotrypsin (chymo). ($n = 3–7$ for **a–d**.)

SNO from IOVs at pH 8 than at pH 7) (Fig. 1d). In addition, incorporation of NO was inhibited substantially by prior modification of IOV thiols with the organic mercurial *p*-chloromercuriphenylsulphonic acid (PCMPs) (a compound that does not affect SNO-Hb binding to the membrane) (Fig. 1d). Together, these results indicate that the NO group is transferred by SNO-Hb to membrane cysteine thiols, which are presumably exposed at the inner surface of the RBC membrane.

The principal interaction of Hb with the RBC membrane is through specific, high-affinity binding to the amino-terminal cytoplasmic domain of the chloride/bicarbonate anion exchange protein AE1 (band 3 protein)^{8–10}. This domain contains two reactive cysteine residues surrounded by amino acids that fit an S-nitrosylation motif¹², which are removed by chymotrypsin^{8,10} and can be

oxidatively crosslinked to the reactive β -chain cysteine in Hb through a disulphide bond¹³. We found that membrane incorporation of NO groups was greatly inhibited after mild digestion of IOVs with chymotrypsin to remove the cytoplasmic N-terminal domain of AE1 (Fig. 1d). In addition, we observed that both SNO-Hb binding and formation of membrane SNO were greatly reduced in IOVs treated with maleimide (data not shown), which has been shown to bind to a lysine residue in the anion exchange domain and disrupt AE1 function¹⁴. Finally, we found that after oxidative crosslinking of SNO-Hb bound to IOV membranes¹³, the Hb β -chain was disulphide-linked selectively to AE1 (Fig. 2h). Therefore, transnitrosylation of a vicinal thiol in the cytoplasmic domain of AE1 is a strong candidate mechanism for transfer of the NO group from the β -chain Cys 93 of SNO-Hb to the RBC membrane.

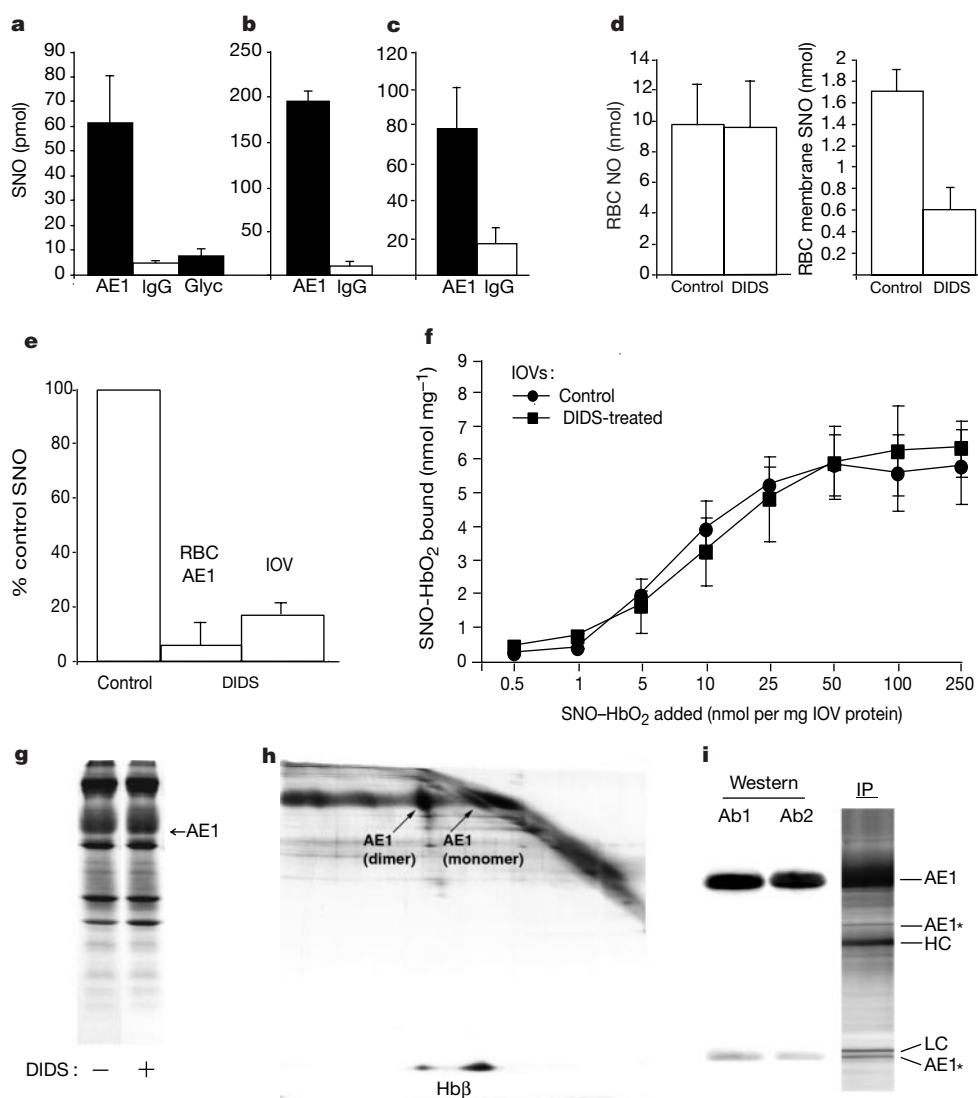


Figure 2 AE1 is S-nitrosylated by SNO-Hb in intact RBCs and IOVs. **a–c**, SNO content of immunoprecipitates derived from IOVs incubated with free (**a**) or Sepharose-bound SNO-Hb (**b**) or from membrane extracts of RBCs treated with NO (NO:haem ratio 1:250) (**c**). Monoclonal antibodies specific for AE1 or glycophorin (Glyc), or a nonspecific mouse IgG were used. **d–g**, Specific inhibition of transnitrosylation by the AE1 inhibitor DIDS (0.1 mM). **d**, Pretreatment with DIDS does not decrease total NO (SNO plus FeNO) incorporated by RBCs (NO:haem ratio 1:250), but substantially decreases membrane SNO content. (Samples were derived from $\sim 2.5 \times 10^9$ RBCs and contained ~ 4 nmol of AE1.) **e**, SNO content is greatly reduced with respect to DIDS-free controls, both in immunoprecipitates of AE1 from membrane extracts of RBCs treated with DIDS before exposure to NO (as in **c**) and in extracts of IOVs derived from DIDS-treated RBCs and

incubated with free SNO-Hb (as in Fig. 1d) ($P < 0.05$). **f**, SNO-Hb binds with equal affinity to IOVs derived from native or DIDS-treated RBCs ($n = 3–8$ for **a–f**). **g**, DIDS does not directly reduce reactivity of RBC membrane thiols as assessed by alkylation with ¹⁴C-iodoacetamide. AE1 is indicated. **h**, The β -chain of Hb (Hb β) is disulphide-coupled selectively to AE1 after incubating IOVs with SNO-Hb and crosslinking by oxidative catalysis. **i**, Both monoclonal antibodies used in immunoprecipitation (Ab1 and Ab2) are specific for AE1 on western blots, and a silver-stained SDS-PAGE gel (IP) shows that AE1 is the predominant constituent of immunoprecipitates of RBC membrane extracts. HC and LC denote heavy and light chains of IgG. Both antibodies identify full-length AE1 (~ 100 K) as well as the characteristic proteolytic N-terminal fragments of AE1 (asterisk) including peptides of ~ 60 K and 20 K (refs 10, 25, 26).

To assess this mechanism further, we incubated IOVs with free or Sepharose-bound SNO-Hb and also treated RBCs with NO (NO:haem ratio 1:250), and then prepared TX100 or NP-40 extracts of IOVs and RBC membranes, immunoprecipitated AE1 with specific monoclonal antibodies (see Fig. 2i and Methods), and measured NO in immunoprecipitates. With standardized quantities of antibody (AE1 in excess), immunoprecipitates generated from extracts of either RBC membranes or IOVs incubated with free SNO-Hb contained about 60–80 pmol of NO, all of which was present as SNO (Fig. 2a, c). Levels of SNO were substantially higher in immunoprecipitates of AE1 from IOVs incubated with immobilized rather than free SNO-Hb (Fig. 2b). SNO was detected at negligible levels (Fig. 2a–c) in immunoprecipitates generated with nonspecific mouse IgG or with a monoclonal antibody to glycoporphin, an abundant AE1-associated RBC membrane protein which also binds Hb¹⁰.

Further evidence for a central role of AE1 in transfer of the NO moiety from SNO-Hb to the RBC membrane was provided by analysing the effects on transnitrosylation of pretreatment of RBCs with the disulphonic stilbene derivative 4,4'-diisothiocyanatostilbene-2,2'-disulphonic acid (DIDS). Stilbene-disulphonates have been used extensively as specific inhibitors of electroneutral anion

exchange and thus of AE1 function in RBCs¹⁵. DIDS acts on AE1 through covalent modification of identified lysine residues in the anion transporter^{15,16}, and this modification induces changes in the conformation of AE1 which can be detected not only in the carboxy-terminal bilayer-spanning domain, but also in the N-terminal cytoplasmic domain^{17–19}.

Incubation of intact RBCs with DIDS before exposure to NO (NO:haem ratio 1:250) did not reduce the net formation of NO-liganded Hb, but greatly decreased the amount of SNO detected in the membrane fraction (Fig. 2d). Furthermore, SNO content was reduced substantially in immunoprecipitates of AE1 from membrane extracts of RBCs treated with DIDS before exposure to NO (Fig. 2e). Transfer of NO groups from SNO-Hb to IOVs prepared from DIDS-treated RBCs was reduced significantly—to a level comparable to that seen after modification of IOV thiols with PCMPs or after chymotryptic digestion of IOV proteins (Fig. 2e). Evidence that DIDS did not affect SNO-Hb binding *per se*, but rather specifically altered the interaction with AE1 required for transnitrosylation, was provided by the finding that SNO-Hb bound with equal affinity to IOVs derived from native or DIDS-treated RBCs (Fig. 2f). In addition, we determined that DIDS did not modify thiols directly, by comparing alkylation with ¹⁴C-iodoacetamide of proteins from native and DIDS-modified IOVs (Fig. 2g). Thus, inhibition of transnitrosylation by DIDS supports the conclusion that AE1 is the principal acceptor of NO groups transferred from SNO-Hb to the RBC membrane.

If RBCs in the systemic circulation transform endothelial-derived NO into SNO and then export physiological amounts of this activity to dilate blood vessels in a process regulated by p_{O_2} (refs 1, 2, 5), and if that process is dependent at least in part on the transnitrosylative mechanism described by our results, then RBCs exposed *in vitro* to low-level (nanomolar) NO should be capable of promoting oxygen-regulated relaxation of arterial smooth muscle, and release of this vasodilatory activity should be inhibited by DIDS treatment (that interferes with NO group transfer from SNO-Hb to the RBC membrane) and perhaps by other AE1 inhibitors.

In bioassays at controlled p_{O_2} (refs 2, 5), we first determined that S-nitrosylated IOVs (containing nanomolar concentrations of

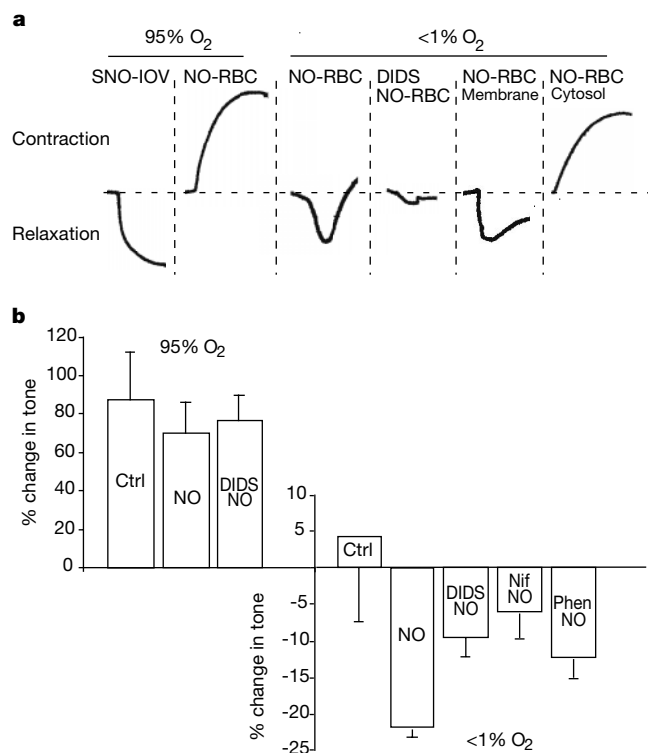


Figure 3 Regulation of aortic tone by RBC consumption and release of NO-related bioactivity. **a**, Representative polygraph traces display the tension developed by aortic ring segments in bioassay medium of specified p_{O_2} . At 95% O_2 , relaxation follows addition of SNO-IOVs (60 nM final SNO concentration), whereas addition of NO-treated RBCs elicits contraction. At <1% O_2 , addition of NO-treated RBCs (containing 60 nM SNO) elicits relaxation, whereas addition of RBCs treated with DIDS before exposure to NO has little effect. Addition of membranes (containing ~60 nM SNO) derived from NO-treated RBCs produces relaxation, whereas an equivalent aliquot of cytosol (containing ~200 nM FeNO) elicits contraction. **b**, Summary of bioassay results ($n = 5–11$) displaying changes in aortic tension elicited at 95% O_2 or less than 1% O_2 by addition of RBCs previously exposed to physiological amounts of NO (NO:haem ratio 1:250; final SNO concentration of 60 nM), RBCs treated with 100 μ M DIDS, niflumic acid (Nif) or phenylglyoxal (Phen) before exposure to NO, and control RBCs deoxygenated and reoxygenated as for NO treatment. DIDS-NO, Nif-NO and Phen-NO at <1% p_{O_2} , $P < 0.05$ versus NO; DIDS-NO and Nif-NO at <1% p_{O_2} , not significant versus control, Ctrl.

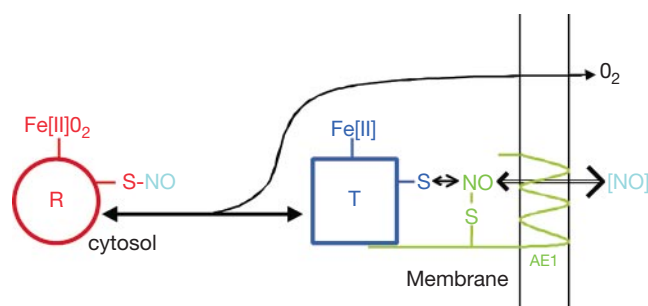


Figure 4 The pathway for export from the RBC of NO-related bioactivity through transfer of NO groups from β -chain Cys 93 of Hb at the membrane-cytosol interface. The cytoplasmic domain of AE1 preferentially binds T-state Hb and thus promotes the R-state to T-state transition of SNO-Hb, which effects transfer of NO groups to the membrane. Therefore, transnitrosylation of vicinal, cysteine thiols in AE1 is a preferred outcome of SNO-Hb deoxygenation in the physiological O_2 gradient. SNO bioactivity ([NO]) is then exported from this site (the role of the anion-exchange domain of AE1 in transmembrane transport is speculative). A juxtamembrane zone of Hb may create a barrier to RBC consumption of extracellular NO (refs 27, 28), which would be strengthened at low p_{O_2} by the preferential association with AE1 of T-state Hb in which cooperativity of NO binding is also lost⁴, thereby decreasing the rate of NO uptake⁴. Thus, the transition from high to low p_{O_2} would be transduced by the interaction of Hb and AE1 into increased export by RBCs of NO-related bioactivity and decreased scavenging of endothelial-derived NO. The fate of NO groups that are in equilibrium between cytosolic SNO-haemoglobin and SNO-glutathione^{1–5} is not considered here.

SNO) can relax rabbit aortic rings and thus that physiological quantities of membrane-associated SNO can convey bioactivity (Fig. 3a). In bioassay medium at 95% O₂, addition of RBCs (bath haematocrit 0.4%; 80 μ M haem) previously exposed to 400 nM NO (NO:haem ratio 1:250) elicited contraction of aortic rings, which was similar in magnitude to that produced by RBCs treated with DIDS before exposure to NO and by control RBCs (which had undergone deoxygenation/reoxygenation as for NO treatment) (Fig. 3). This observation is consistent with studies showing that free oxyHb and SNO-Hb increase aortic smooth muscle tone at high p_{O_2} (refs 1, 2, 5, 6), which can be ascribed to scavenging by R-structured Hb of NO produced by basal activity of endothelial NO synthase^{5,6}.

In contrast, aortic tone in medium at less than 1% O₂ to simulate tissue p_{O_2} was relaxed significantly ($21.7 \pm 1\%$; $n = 11$) by addition of RBCs previously exposed to NO (60 nM final SNO concentration, as with S-nitrosylated IOVs) (Fig. 3). Relaxation of this magnitude in arterioles would produce about a 200% increase in blood flow. It is important to recognize that NO was added to RBCs many minutes before they were used, and thus that RBCs preserved NO bioactivity to release it on demand. In addition, it should be noted that relaxation elicited at low p_{O_2} by SNO-RBCs (unlike SNO-IOVs) was followed by contraction once vasodilatory activity was depleted (data not shown). When NO-treated RBCs (NO:haem ratio 1:250) were separated into membrane and cytosolic fractions, aliquots of membrane (containing ~ 60 nM SNO) elicited relaxation at low p_{O_2} similar to that produced by intact, NO-treated RBCs, whereas equal portions of the cytosolic fraction (containing ~ 200 nM HbFeNO) elicited contraction (Fig. 3a). Thus in this preparation, generation by RBCs of Hb-derived NO bioactivity requires both low p_{O_2} and the interaction of Hb with the RBC membrane. RBCs treated with DIDS before exposure to NO produced substantially less relaxation ($9.4 \pm 2.8\%$, $n = 5$) than RBCs exposed to NO alone, to the extent that the average response after DIDS did not differ significantly from that evoked by control RBCs at less than 1% O₂ (Fig. 3). Furthermore, the AE1 inhibitors niflumic acid and phenylglyoxal^{15,20} produced abrogated relaxations at low p_{O_2} of $6.5 \pm 3.3\%$ ($n = 3$) and $12.3 \pm 2.6\%$ ($n = 3$), respectively (Fig. 3b). These results show that transition from high to low p_{O_2} evokes the export of vasodilatory NO bioactivity from RBCs (containing physiological amounts of NO), and that release is inhibited by disrupting the transnitrosylative transfer of NO groups from SNO-Hb to cysteine thiols in the cytoplasmic domain of AE1.

In addition, it is notable that contractions elicited by control RBCs were consistently weaker at less than 1% O₂ than at 95% O₂ (Fig. 3b). This observation contrasts with the increase in tone elicited by free Hb, which is not reduced (or even enhanced) at low p_{O_2} (refs 2, 6), and suggests that net import of extracellular (endothelial-derived) NO to cytosolic Hb is inhibited under those conditions where conjoint delivery of O₂ and NO bioactivity is called for (see Fig. 4).

Our findings indicate that the interaction of Hb with AE1 at the membrane (Fig. 4) must be incorporated in the scheme that describes the reactions governing the fate of NO in the RBC. The site of Hb binding in the cytoplasmic domain of AE1 has been localized to the polyanionic N terminus, and analysis of co-crystals of Hb and an AE1 N-terminal peptide has shown that this stretch of acidic residues inserts into the 2,3-diphosphoglycerate-binding pocket formed between β -globin subunits of tetrameric Hb (ref. 9). Consistent with this binding mechanism, AE1 binds with higher affinity to deoxyHb than to oxyHb (in which the β -cleft is occluded)^{9,21}, and consistent with the requirements of thermodynamic linkage⁵, the oxygen affinity of Hb is reduced in the presence of isolated AE1 cytoplasmic domain⁹. AE1 is thus both an allosteric regulator that promotes Hb T-state and an acceptor of NO groups transferred by SNO-Hb upon R-state to T-state transition. Therefore, S-nitrosylation of AE1 is a preferred outcome of SNO-Hb

deoxygenation at the membrane-cytosol interface (Fig. 4). AE1 is present at $\sim 10^6$ copies per RBC¹⁰, ~ 10 -fold excess over SNO-Hb in arterial RBCs^{1,2}, which is sufficient for the proposition that transfer to AE1 constitutes a main route of NO trafficking as the RBC transits the physiological p_{O_2} gradient².

Although our findings indicate that DIDS-induced inhibition of export of NO-related bioactivity can be accounted for by reduced transfer of NO groups from SNO-Hb to cysteine thiols in the cytoplasmic domain of AE1, inhibition by different classes of AE1 inhibitors (Fig. 3) suggests that a role for the anion exchange function of AE1 in transmembrane transport of NO bioactivity should also be considered. The anion selectivity of AE1 is not strict: erythrocyte AE1 transports several NO_x species²²⁻²⁴, and other NO-related anions may be substrates. Thus, although RBCs can release bioactive S-nitrosylated low molecular mass thiols¹, it is an appealing possibility that NO bioequivalents cross the RBC membrane, at least in part, by means of AE1-mediated transport of an NO congener or NO_x derived in the immediate juxtamembrane locale.

Our findings demonstrate the physiological protein-protein transfer of NO groups, and signify that transnitrosylation is a mechanism of post-translational protein modification involved in the transport and targeting of NO bioactivity in signal transduction¹². Our results also provide a new perspective on the transmembrane translocation of NO bioactivity, which both emphasizes segregation at or in the membrane of reaction pathways for transfer of NO equivalents and raises the possibility of regulation by NO of diverse membrane functions. Finally, our results highlight the potential but largely unexplored role of Hb-membrane interaction in the pathogenesis of the thrombotic diatheses, ischaemic syndromes, oxidative disorders and hypertensive states that are associated with altered haematocrit, haemoglobinopathies and RBC membrane defects. □

Methods

Reactions of NO/SNO with RBCs and IOVs

RBCs derived from fresh venous blood and suspended in oxygen-purged PBS were exposed in gas-tight vials to NO, added as aliquots of oxygen-purged NO-saturated PBS to yield the specified NO:haem ratios, based on spectrophotometric assessment of haem content. Exposure to NO was for 5 min followed by rinses in PBS at 21% O₂. RBCs were lysed in 10 mM MOPS buffer, 0.1 mM diethylene triaminepentaacetic acid, pH 7, followed by centrifugation at 20,000g for 10 min to produce membrane and cytosolic fractions, which were solubilized in 1% TX100. Quantification of NO groups was by photolysis/chemiluminescence⁶. SNO content was measured as the amount of NO removed by treating sample aliquots for 5 min with HgCl₂ at a fivefold molar excess over estimated thiol content^{3,4,6}. IOVs were prepared from outdated RBCs as described¹¹ and stripped of peripheral membrane-bound proteins by washing for 60 min in 0.5 M NaCl/100 mM Tris, pH 8. Chymotrypsin treatment of IOVs was for 15 min at pH 7 and 37 °C with 5 U enzyme per mg IOV protein. For transnitrosylation, IOVs were incubated with SNO-Hb (50 nmol SNO-Hb per mg IOV protein, about 10-fold excess over AE1) for 15 min at pH 7 and 37 °C. SNO-Hb was prepared from free or immobilized HbAo (Apex) by selective S-nitrosylation of β -chain Cys 93 with S-nitrosocysteine^{2,6}. Hb was immobilized on cyanogen bromide-activated Sepharose 4b.

Immunoprecipitation, DIDS-modification, and crosslinking

Immunoprecipitation was done by incubation of TX100 or NP-40 extracts with antibodies followed by protein G/Sepharose. We used two extensively characterized monoclonal anti-AE1 antibodies (Sigma, clone BIII-136, and TE10, kindly provided by M. Telen) that recognize epitopes in the cytoplasmic domain of AE1 (refs 25, 26) and that produced indistinguishable results (see Fig. 2i). Immune complexes were eluted with 100 mM glycine at pH 2.5–3.0, and eluates were adjusted to pH 7.4 before photolysis/chemiluminescence. Efficiency of immunoprecipitation was confirmed by western blotting. Binding of SNO-Hb to IOVs was assessed in 10 mM MOPS buffer, pH 7, at room temperature and 21% O₂. After 15 min, IOVs were solubilized in 1% TX100, and haem content was assessed spectrophotometrically. When used, RBCs were treated for 60 min with DIDS (0.1 mM; calculated 10-fold molar excess over AE1), and then washed thoroughly to remove unreacted DIDS. RBCs were then exposed to NO or used to prepare DIDS-modified IOVs, as above. To identify RBC membrane proteins that could be coupled to bound SNO-Hb by a disulphide linkage (vicinal thiols), IOVs were incubated with SNO-Hb, rinsed to remove unbound SNO-Hb, and crosslinked by oxidative catalysis with Cu²⁺-*o*-phenanthroline¹³. IOV-Hb complexes were analysed by two-dimensional SDS-PAGE, with non-reducing first dimension and reducing second dimension¹³, followed by silver staining. The position of AE1 was identified by western

blotting of gels run in parallel, and Hb(β) and Hb(α) were discriminated by their differential electrophoretic mobility.

Bioassay

Rabbit aortic rings were suspended in Krebs-bicarbonate buffer at 37°C, bubbled continuously with either 95% O₂/5% CO₂ or 95% argon/5% CO₂ (measured O₂ less than 1%)^{1,2}. Resting tension was maintained at a standard 2 g and active tension was induced by phenylephrine. NO-treated and control RBCs were washed and resuspended in PBS at 50% haematocrit, and added to individual 25-ml baths as 0.2-ml aliquots to yield a bath haematocrit of 0.4%. RBCs were treated with AE1 inhibitors (DIDS, niflumic acid or phenylglyoxal; 0.1 mM in all cases) for 60 min before washing and exposure to NO.

Received 20 April; accepted 6 November 2000.

- Jia, L., Bonaventura, C., Bonaventura, J. & Stamler, J. S. S-nitrosohaemoglobin: a dynamic activity of blood involved in vascular control. *Nature* **380**, 221–226 (1996).
- Stamler, J. S. *et al.* Blood flow regulation by S-nitrosohaemoglobin is controlled by the physiological oxygen gradient. *Science* **276**, 2034–2037 (1997).
- Gow, A. J. & Stamler, J. S. Reactions between nitric oxide and haemoglobin under physiological conditions. *Nature* **391**, 169–173 (1998).
- Gow, A. J., Luchsinger, B. P., Pawloski, J. R., Singel, D. J. & Stamler, J. S. The oxyhemoglobin reaction of nitric oxide. *Proc. Natl Acad. Sci. USA* **96**, 9027–9032 (1999).
- McMahon, T. J., Stone, A. E., Bonaventura, J., Singel, D. J. & Stamler, J. S. Functional coupling of oxygen-binding and vasoactivity in S-nitrosohaemoglobin. *J. Biol. Chem.* **275**, 16738–16745 (2000).
- McMahon, T. J. & Stamler, J. S. Concerted nitric oxide/oxygen delivery by hemoglobin. *Methods Enzymol.* **301**, 99–114 (1999).
- Pawloski, J. R., Swaminathan, R. V. & Stamler, J. S. Cell-free and erythrocytic S-nitrosohaemoglobin inhibits human platelet aggregation. *Circulation* **97**, 263–267 (1998).
- Rauenbuehler, P. B., Cordes, K. A. & Salhany, J. M. Identification of the hemoglobin binding sites on the inner surface of the erythrocyte membrane. *Biochim. Biophys. Acta* **692**, 361–370 (1982).
- Waldner, J. A. *et al.* The interaction of hemoglobin with the cytoplasmic domain of band 3 of the human erythrocyte membrane. *J. Biol. Chem.* **259**, 10238–10246 (1984).
- Low, P. S. Structure and function of the cytoplasmic domain of band 3: center of erythrocyte membrane-peripheral protein interactions. *Biochim. Biophys. Acta* **864**, 145–167 (1986).
- Kondo, T. Preparation of microcapsules from human erythrocytes: use in transport experiments of glutathione and its S-conjugate. *Methods Enzymol.* **171**, 217–225 (1989).
- Stamler, J. S., Toone, E. J., Lipton, S. A. & Sucher, N. J. (S)NO signals: translocation, regulation, and a consensus motif. *Neuron* **18**, 691–696 (1997).
- Sayare, M. & Fikiet, M. Cross-linking of hemoglobin to the cytoplasmic surface of human erythrocyte membranes. Identification of band 3 as a site for hemoglobin binding in Cu²⁺-*o*-phenanthroline catalyzed cross-linking. *J. Biol. Chem.* **256**, 13152–13158 (1981).
- Liu, S. Q. & Knauf, P. A. Lys-430, site of irreversible inhibition of band 3 Cl⁻ flux by eosin-5-maleimide, is not at the transport site. *Am. J. Physiol.* **264**, C1155–C1164 (1993).
- Falke, J. J. & Chan, S. I. Molecular mechanisms of band 3 inhibitors. 1. Transport site inhibitors. *Biochemistry* **25**, 7888–7894 (1986).
- Okubo, K., Kang, D., Hamaaki, N. & Jennings, M. L. Red blood cell band 3. Lysine 539 and lysine 851 react with the same H2DIDS (4,4'-diisothiocyanodihydrostilbene-2,2'-disulfonic acid) molecule. *J. Biol. Chem.* **269**, 1918–1926 (1994).
- Salhany, J. M., Cordes, K. A. & Gaines, E. D. Light-scattering measurements of hemoglobin binding to the erythrocyte membrane. Evidence for transmembrane effects related to a disulfonic stilbene binding to band 3. *Biochemistry* **19**, 1447–1454 (1980).
- Hsu, L. & Morrison, M. The interaction of human erythrocyte band 3 with cytoskeletal proteins. *Archiv. Biochem. Biophys.* **227**, 31–38 (1983).
- Macara, I. G., Kuo, S. & Cantley, L. C. Evidence that inhibitors of anion exchange induce a transmembrane conformational change in band 3. *J. Biol. Chem.* **258**, 1785–1792 (1983).
- Falke, J. J. & Chan, S. I. Molecular mechanisms of band 3 inhibitors. 3. Translocation inhibitors. *Biochemistry* **25**, 7899–7906 (1986).
- Chetrite, G. & Cassoly, R. Affinity of hemoglobin for the cytoplasmic fragment of human erythrocyte membrane band 3. Equilibrium measurements at physiological pH using matrix-bound proteins: the effects of ionic strength, deoxygenation and of 2,3-diphosphoglycerate. *J. Mol. Biol.* **185**, 639–644 (1985).
- Galanter, W. L. & Labotka, R. J. The binding of nitrate to the human anion exchange protein (AE1) studied with ¹⁴C nuclear magnetic resonance. *Biochim. Biophys. Acta* **1079**, 146–151 (1991).
- Shingles, R., Roh, M. H. & McCarty, R. E. Direct measurement of nitrite transport across erythrocyte membrane vesicles using the fluorescent probe, 6-methoxy-N-(3-sulfolpropyl) quinolinium. *J. Bioenerg. Biomembr.* **29**, 611–616 (1997).
- Soszynski, M. & Bartosz, G. Penetration of erythrocyte membrane by peroxynitrite: participation of the anion exchange protein. *Biochem. Mol. Biol. Int.* **43**, 319–325 (1997).
- Czerwinski, M. *et al.* Degradation of the human erythrocyte membrane band 3 studied with the monoclonal antibody directed against an epitope on the cytoplasmic fragment of band 3. *Eur. J. Biochem.* **174**, 647–654 (1988).
- Telen, M. J., Scarsie, R. M. & Haynes, B. Human erythrocyte antigens. III. Characterization of a panel of murine monoclonal antibodies that react with human erythrocyte and erythroid precursor membranes. *Vox Sang.* **52**, 236–243 (1987).
- Liu, X. *et al.* Diffusion-limited reaction of free nitric oxide with erythrocytes. *J. Biol. Chem.* **273**, 18709–18713 (1998).
- Vaughn, M. W., Huang, K.-T., Kuo, L. & Liao, J. C. Erythrocytes possess an intrinsic barrier to nitric oxide consumption. *J. Biol. Chem.* **275**, 2342–2348 (2000).

Acknowledgements

We are grateful to M. Telen for providing antibody to AE1 and for advice. This work was supported by an AHA postdoctoral fellowship to J.R.P.

Correspondence and requests for materials should be addressed to J.S.S. (e-mail: STAM1001@mc.duke.edu).

Scabrous complexes with Notch to mediate boundary formation

Patricia A. Powell^{†‡}, Cedric Wesley^{§‡}, Susan Spencer^{*} & Ross L. Cagan^{*}

^{*} Department of Molecular Biology and Pharmacology, Washington University School of Medicine, 660 South Euclid Avenue, Campus Box 8103, St. Louis, Missouri 63110, USA

[§] Laboratory of Genetics, The Rockefeller University, 1230 York Avenue, York, New York 10021, USA

[†] These authors contributed equally to this work

The mechanisms that establish and sharpen pattern across epithelia are poorly understood. In the developing nervous system, the first pattern elements appear as 'proneural clusters'. In the morphogenetic furrow of the immature *Drosophila* retina proneural clusters emerge in a wave as a patterned array of 6–10-cell groups, which are recognizable by expression of Atonal, a basic helix–loop–helix transcription factor that is required to establish and pattern the first cell fate^{1–3}. The establishment and subsequent patterning of Atonal expression requires activity of the signalling transmembrane receptor Notch^{2,4}. Here we present *in vivo* and biochemical evidence that the secreted protein Scabrous associates with Notch, and can stabilize Notch protein at the surface. The result is a regulation of Notch activity that sharpens proneural cluster boundaries and ensures establishment of single pioneer neurons.

In the morphogenetic furrow, Atonal's expression can be divided into four steps^{1–3} (Fig. 1a). First, it is expressed as a broad, unpatterned stripe; second, expression is then upregulated into evenly spaced proneural clusters; third, in these proneural clusters, a 2–3-cell 'R8 equivalence group' emerges; and last, expression narrows to identify a single cell in this group as the R8 photoreceptor neuron, the first cell type of the developing retina. In each step, as cells lose Atonal expression they concurrently gain expression of negative regulators, such as members of the E(spl) (Enhancer of Split) complex^{1,2} (Fig. 1a). Expression of E(spl) initially requires the presence of Atonal³, and is subsequently amplified by the Notch signalling pathway to downregulate proneural bHLH expression and function^{4,5}. E(spl), therefore, represents one reporter of Notch activity.

Patterning of Atonal and E(spl) expression requires normal activity of Scabrous, a secreted fibrinogen-related protein with a potential for association with components of the extracellular matrix^{6–8}. In the retina, Scabrous protein first appears in the proneural clusters, mirroring Atonal expression by narrowing to the R8 equivalence group, and eventually R8 alone. This expression is dependent on Atonal activity⁹, which indicates that the *scabrous* locus may be a direct target of Atonal.

Genotypically null *sca*^{BP2} proneural clusters are poorly spaced with poorly defined borders (Fig. 1b)^{10,11}. Broadened E(spl) expression throughout much of the proneural cluster region is one potential cause of this imprecision (Fig. 1b), suggesting that Notch activity is altered in *sca*^{BP2} mutants. These observations suggest that initial broad, low-level Atonal expression activates broad, low-level E(spl) expression, and that Scabrous is required to refine the complementary Atonal and E(spl) expression in the proneural cluster region—events also associated with Notch activity.

A 1-h pulse of ectopic Scabrous resulted in rapid loss of Atonal within 2 h (Fig. 1d); E(spl) showed low, diffuse expression (Fig. 1d) that was lost within 4 h (data not shown). This ectopic expression of

[‡] Present address: National Science Foundation, 4201 Wilson Blvd., Arlington, Virginia 22230, USA.

Scabrous leads to aberrant patterning of R8s in a manner similar to that of the phenotypes observed in *scabrous* loss-of-function alleles¹¹. Corresponding wild-type heat shock had no effect (data not shown). The loss of the initial broad stripe of Atonal, a *Notch*-dependent step¹², suggests that ectopic Scabrous can lead to a disruption of *Notch* function, a result consistent with overexpression studies in the *Drosophila* wing¹³.

Subsequent resolution of Atonal expression in the equivalence group, also a *Notch*-dependent step^{10,14,15}, is delayed in *sca*^{BP2} mutants (Fig. 1b). Delay of Atonal resolution can result in ectopic R8s (ref. 1); indeed, 40% of *sca*^{BP2} ommatidia ($n = 200$) contain two or three R8s derived from the R8 equivalence group (Fig. 1c). R8s were observed to be both adjacent and non-adjacent, indicating that selection may be stochastic.

These results suggest that Scabrous influences the establishment or maintenance of boundaries of *Notch* activity during both proneural-cluster and R8 equivalence-group maturation. To investigate the nature of this influence, transient expression of full-length Notch protein (N350) in *Drosophila* S2 tissue-culture cells was examined by western analysis (Fig. 2a; $n = 14$ experiments). Transiently induced N350 was depleted gradually over 7 h; this loss was accelerated in the presence of Delta (Fig. 2a). In contrast, addition of Scabrous led to an accumulation of N350 and smaller Notch fragments, even in the presence of Delta (Fig. 2a). Accumulation of Notch was not due to induction of the endogenous *Notch* gene, which is rearranged in S2 cells (ref. 16; and C.S.W., unpub-

lished observations). Cell-surface biotinylation analysis indicated that the presence of Scabrous significantly stabilized N350 at the cell surface (Fig. 2b); however, the functional nature of this stabilization remains unknown.

Genetic and histological studies have suggested a function for Scabrous as a secreted Notch ligand^{10,17}. However, attempts to show this interaction biochemically have been unsuccessful¹⁸. We returned to this issue using two independent assays. Notch is expressed throughout embryogenesis^{19,20}, whereas Scabrous protein was detected in young (2.5–6.5 h) but not more mature (12.5–16.5 h) embryos (Fig. 2c). Immunoprecipitation with a mouse polyclonal anti-Scabrous antibody led to a complex with a relative molecular mass of 300,000 (M_r 300K) that contained both Scabrous and Notch protein (Fig. 2c). The size and composition of the 300K complex indicates that it may include a truncated form of Notch previously characterized *in vivo*²¹. As controls, crosslinked complexes recovered by an anti-Delta antibody included Notch, but complexes recovered by an anti-Fasciclin III antibody did not (data not shown).

To investigate binding *in vivo*, we used the fact that diffusible ligands can be stabilized specifically in cells expressing their natural receptor²². In the midpupal eye, Notch is expressed exclusively in a set of interommatidial cells, the secondary and tertiary pigment cells (Fig. 3a)²⁰. Scabrous is not expressed in the retina at this developmental stage (Fig. 3b). Transient heat-shock induction of *hs-scabrous* pupae resulted in a ubiquitous pulse of Scabrous protein

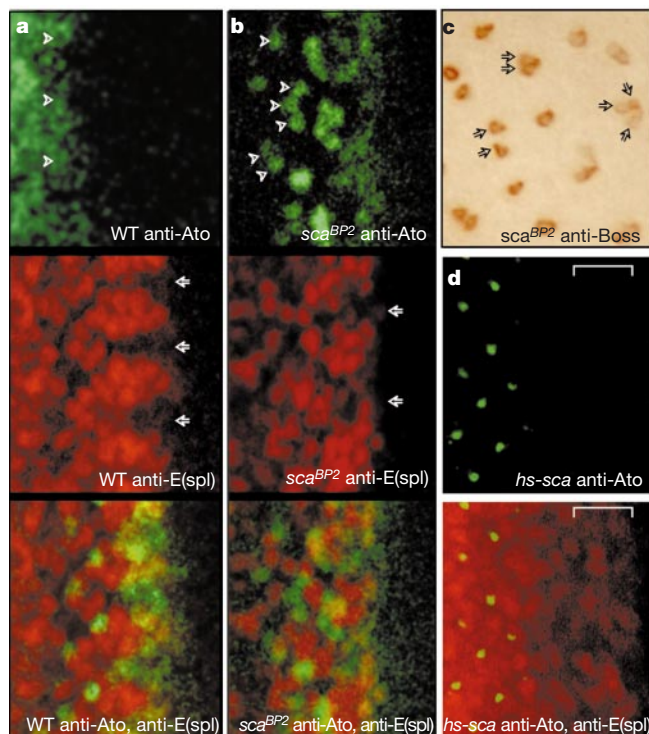


Figure 1 Altering Scabrous activity. Anterior is toward the right. **a**, Atonal and E(spl) expression in wild-type (WT) morphogenetic furrow. Top, Atonal expression evolves from low and diffuse to high in regularly spaced proneural clusters and eventually R8s (arrowheads). Middle, E(spl) (m8, m7, m6 and m3) extends between, but not in (arrows), proneural clusters. **b**, Atonal and E(spl) expression in *sca*^{BP2} morphogenetic furrow. Top, Atonal expression is low and diffuse; proneural clusters are shorter with poorly delineated boundaries. Expression is later retained in 2–3 cells (arrowheads). Middle, E(spl) extends into proneural cluster regions (arrows). Poorly spaced *sca*^{BP2} ommatidia often contain several R8s (marked with Boss; arrows) **(c)**. **d**, Effect of ectopic Scabrous (*hs-scabrous*) on Atonal and E(spl) expression. Top, *hs-scabrous* blocks Atonal expression in morphogenetic furrow (bracket). R8s are labelled with Boss. Bottom, E(spl) is unpatterned and diffuse.

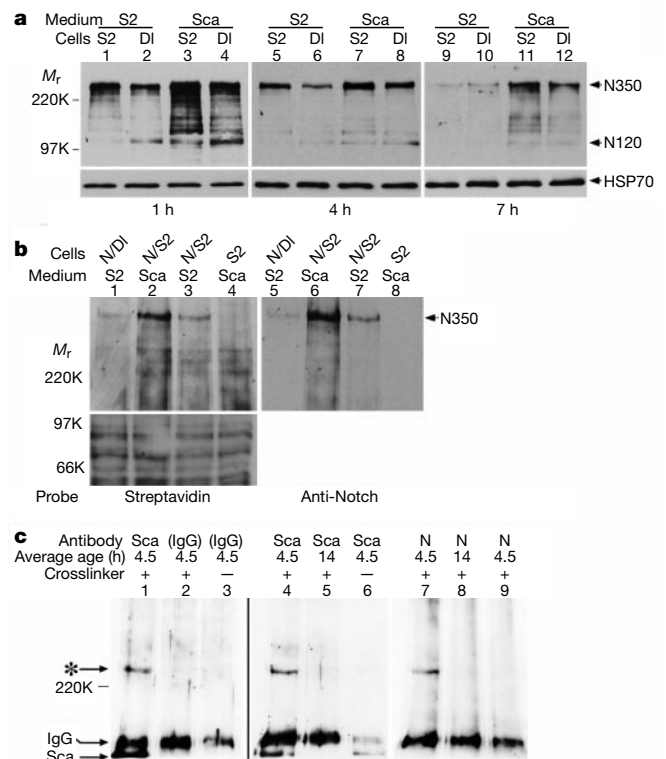


Figure 2 Scabrous stabilizes Notch and forms a complex *in vivo*. **a**, Western analysis of S2 tissue-culture cells. S2-Notch cells treated with Scabrous for 1, 4 or 7 h have higher N350 levels even in presence of Delta. N120 is a Notch intracellular fragment. Conditioned media: S2, S2-hsSca. Cell lines: S2, S2-Notch, S2-Dl. **b**, Cell-surface biotinylation analysis. More N350 molecules are biotinylated in the presence of Scabrous. Other cell-surface molecules are biotinylated at similar levels (below lanes 1–4). Lanes 5–8 are the same as lanes 1–4, but reprobed with Notch antibody to detect N350. **c**, Scabrous antibody immunoprecipitates ~300K complex (asterisk) containing Scabrous and Notch (lanes 1, 4, 7) during stages of high Scabrous expression (lanes 4, 5).

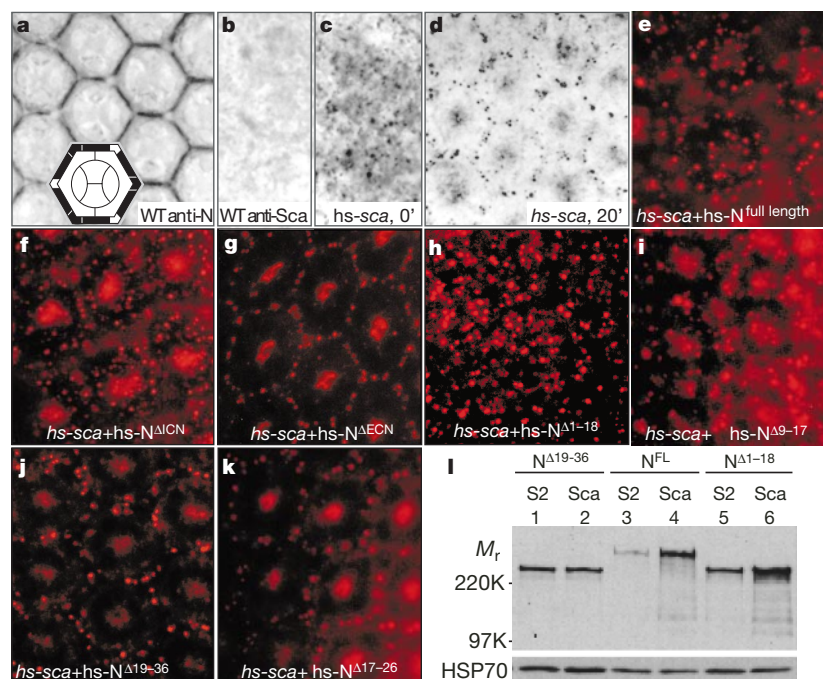


Figure 3 Scabrous and Notch colocalize *in vivo*. Apical surface of 38-h pupal retinæ. **a**, Notch expression in interommatidial cells. Notch is expressed exclusively in secondary/tertiary pigment cells (shaded; inset). **b**, Scabrous is not expressed. Shifting *hs-scabrous* pupae to 37 °C results in broad Scabrous expression (**c**) that is lost 10 min later (**d**) except in secondary/tertiary pigment cells. The strong staining at the ommatidial centre is

Scabrous that is trapped in photoreceptor rhabdomeres. Full-length Notch, N^{Δ1CN}, N^{Δ9-17} and N^{Δ1-18} (**e, f, h, i**), but not N^{ΔECN}, N^{Δ17-26} or N^{Δ19-36} (**g, j, k**), stabilizes Scabrous in all cells. **l**, After 4 h, addition of Scabrous stabilizes full-length Notch and N^{Δ1-18}, but not Notch^{Δ19-36}.

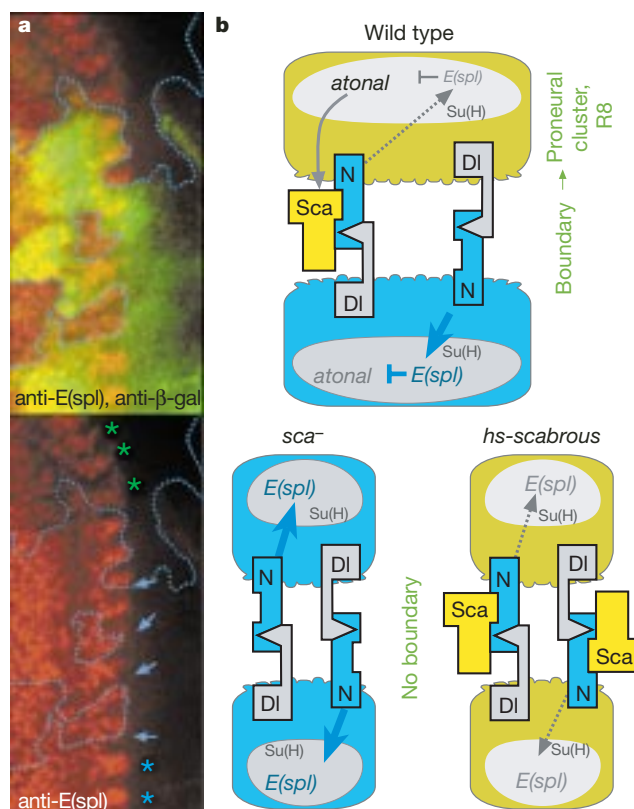


Figure 4 Scabrous and boundaries. **a**, Top, Lack of β-galactosidase expression (cytoplasmic, green) in *sca^{BP2}* clones (dotted lines). Bottom, Within, but near clone edges (blue arrows), *E(spl)* expression (nuclear, red) often re-orientates to match clone boundary. Compare with sharp wild-type (blue asterisks) and diffuse, fully *sca^{BP2}* (green asterisks) boundaries. **b**, Model. Cells expressing high levels of Atonal (proneural cluster,

R8 equivalence group) express Scabrous, reducing their *Notch* activity and *E(spl)* expression (dotted arrow). Neighbouring cells that fail to express Scabrous show high *notch* activity (solid arrow), further reducing Atonal expression. Initial Atonal differences are amplified, stabilizing proneural cluster edges and later ensuring single R8s. Su(H), Suppressor of Hairless.

(Fig. 3c) that was processed, secreted and rapidly lost from most cells in as little as 5 min after the end of heat shock. However, Scabrous was retained in secondary/tertiary cells (Fig. 3d), the same cells that express the Notch receptor. Retention of Scabrous occurred in apically localized vesicles commonly observed with secreted ligands and their receptors, including Notch²³. This cell-type-specific retention suggests that a receptor for Scabrous is found specifically in the secondary/tertiary cells at this stage of development.

We next altered the expression pattern of Notch through an inducible heat-shock promoter. Co-expression of Scabrous and Notch in all retinal cells resulted in the stabilization of Scabrous in all cells (Fig. 3e). Similarly, ubiquitous expression of an inactive form of Notch that lacked its intracellular domain²⁴ ($N^{\Delta ICN}$) was also sufficient to stabilize exogenous Scabrous in all cells (Fig. 3f), indicating that Scabrous may associate with the extracellular domain of Notch. This also indicates that stabilization of Scabrous did not require the signalling activity of Notch. As predicted, a membrane-tethered form of the intracellular domain²⁴ ($N^{\Delta ECN}$) did not retain Scabrous (Fig. 3g).

The extracellular domain of Notch includes 36 epidermal growth factor (EGF)-like repeats. Two demonstrated ligands of Notch, Delta and Serrate, require EGF-like repeats 11 and 12 for binding²⁵. Ubiquitous expression of Notch forms that deleted repeats 1–18 or 9–17 ($N^{\Delta 1-18}$ and $N^{\Delta 9-17}$) was still sufficient to stabilize Scabrous protein in all retinal cells (Figs 3h, i); however, $N^{\Delta 19-36}$ and $N^{\Delta 17-26}$ failed to retain Scabrous protein (Figs 3j, k). Western analysis indicated that $N^{\Delta 9-17}$ and $N^{\Delta 17-26}$ were expressed at similar levels in pupal eye discs after heat shock (data not shown). The inability of $N^{\Delta 19-36}$ or $N^{\Delta 17-26}$ to retain Scabrous suggests that the site of Notch/Scabrous association lies within repeats 19–26. Consistent with this view, Scabrous protein was able to promote stabilization of $N^{\Delta 1-18}$ that was expressed on the surface of S2 cells, but failed to stabilize $N^{\Delta 19-36}$ (Fig. 3l).

Our data support the view^{10,13,17} that Scabrous and Notch form a receptor–ligand pair; however, we cannot rule out the possibility that binding requires other cofactors or adapters. Our results indicating stabilization suggest that Scabrous may act to scaffold Notch protein to the extracellular matrix and downregulate Notch activity in at least some experimental situations¹³ to sharpen local boundaries. Consistent with this view, the sharp boundaries between proneural clusters were lost in *sca*^{BP2} clonal patches (Fig. 4a), indicating that Scabrous may act over a short distance. Notably, when a *sca*^{BP2} clonal patch crossed the morphogenetic furrow, high E(spl) expression typically coincided with the boundary of the clone (14 out of 18 *sca*^{BP2} clonal patches adjacent to a proneural cluster followed its boundary, as compared with 5 out of 14 wild-type control clones).

The observed complexity of their *in vivo* interactions leaves the precise mechanism by which Scabrous regulates Notch at boundaries unresolved. One model is presented in Fig. 4b. Notch signalling establishes boundaries during a variety of epithelial developmental decisions, and similar to Fringe²⁶, Scabrous provides a means by which Notch activity can be altered in the absence of changes in ligand expression, although the tissues and possibly the mechanisms, are different. □

Methods

Heat shocks

hs-scabrous larvae received a 1-h heat pulse at 37 °C, and recovered for 1–4 h. For *in vivo* binding, white prepupae were aged at 25 °C for 37 h, then heat shocked at 37 °C for 2 × 12 min with intervening 90 min recovery at 25 °C, held at 25 °C for 20 min, and iced. We then dissected the pupal eyes directly into PLP fixative¹.

Immunohistochemistry

The antibodies used were anti-Atonal (1:10,000)⁹, anti-E(spl)323 (1:10)²⁷, anti-Notch (1:50)¹⁶ and anti-Scabrous (1:10)¹⁰. We used standard fixation and staining protocols¹.

Confocal images for Fig. 1a and b were captured, and projections created with ImageSpace software (Molecular Dynamics).

Clonal analysis

yw hFLP;FRT42 conD lacZ/CyO flies were crossed with *FRT42 sca*^{BP2}/Bc *Gla* or (as a control) *FRT42* flies. To prevent scoring bias, tracings of the *lacZ* were done in the absence of the E(spl) image. We only scored proneural clusters that lay at least partially outside (but adjacent) to the clones.

Immunoprecipitations

Crosslinked cell surface molecules were immunoprecipitated with a Scabrous antibody¹² and the complexes detected on western blots with the same Scabrous antibody or a Notch antibody²⁴ as described²¹. We reprobed lanes 4–6 in Fig. 2c to produce lanes 7–9.

Cell culture and western analysis

Heat shocked 10⁷ S2 or S2-*hs-scabrous* cells were used to produce 1 ml of conditioned serum-free Shields and Sang M3 medium for treating cells. Cell surfaces were biotinylated for 15 min with 1 mg ml⁻¹ Sulpho-NHS-LC-Biotin (Pierce) in HEPES (pH 8.1)-buffered Chan and Gehring medium (Ashburner). Proteins were extracted, quantified and analysed on 4% or 8% SDS-PAGE as described using anti-Notch or streptavidin-conjugated antibodies^{21,28,29}. Total protein levels in lanes or samples were assessed using an antibody against HSP 70 (Sigma). Specificity of cell-surface biotinylation was checked by non-biotinylation of the intracellular HSP 70.

Received 22 May; accepted 6 November 2000.

- Dokucu, M. E., Zipursky, S. L. & Cagan, R. L. Atonal, rough and the resolution of proneural clusters in the developing *Drosophila* retina. *Development* **122**, 4139–4147 (1996).
- Baker, N. E., Yu, S. & Han, D. Evolution of proneural atonal expression during distinct regulatory phases in the developing *Drosophila* eye. *Curr. Biol.* **6**, 1290–1301 (1996).
- Jarman, A. P., Sun, Y., Jan, L. Y. & Jan, Y. N. Role of the proneural gene, atonal, in formation of *Drosophila* chordotonal organs and photoreceptors. *Development* **121**, 2019–2030 (1995).
- Bailey, A. M. & Posakony, J. W. Suppressor of hairless directly activates transcription of enhancer of split complex genes in response to Notch receptor activity. *Genes Dev.* **9**, 2609–2622 (1995).
- Lecourtis, M. & Schweisguth, F. The neurogenic suppressor of hairless DNA-binding protein mediates the transcriptional activation of the enhancer of split complex genes triggered by Notch signaling. *Genes Dev.* **9**, 2598–2608 (1995).
- Lee, E. C., Hu, X., Yu, S. Y. & Baker, N. E. The scabrous gene encodes a secreted glycoprotein dimer and regulates proneural development in *Drosophila* eyes. *Mol. Cell. Biol.* **16**, 1179–1188 (1996).
- Lee, E. C., Yu, S. Y., Hu, X., Mlodzik, M. & Baker, N. E. Functional analysis of the fibrinogen-related scabrous gene from *Drosophila melanogaster* identifies potential effector and stimulatory protein domains. *Genetics* **150**, 663–673 (1998).
- Mlodzik, M., Baker, N. E. & Rubin, G. M. Isolation and expression of scabrous, a gene regulating neurogenesis in *Drosophila*. *Genes Dev.* **4**, 1848–1861 (1990).
- Jarman, A. P., Grell, E. H., Ackerman, L., Jan, L. Y. & Jan, Y. N. Atonal is the proneural gene for *Drosophila* photoreceptors. *Nature* **369**, 398–400 (1994).
- Baker, N. E., Mlodzik, M. & Rubin, G. M. Spacing differentiation in the developing *Drosophila* eye: a fibrinogen-related lateral inhibitor encoded by scabrous. *Science* **250**, 1370–1377 (1990).
- Ellis, M. C., Weber, U., Wiersdorff, V. & Mlodzik, M. Confrontation of scabrous expressing and non-expressing cells is essential for normal ommatidial spacing in the *Drosophila* eye. *Development* **120**, 1959–1969 (1994).
- Ligoxysakis, P., Yu, S. Y., Delidakis, C. & Baker, N. E. A subset of Notch functions during *Drosophila* eye development require Su(H) and the E(spl) gene complex. *Development* **125**, 2893–2900 (1998).
- Lee, E. C., Yu, S. Y. & Baker, N. E. The scabrous protein can act as an extracellular antagonist of Notch signaling in the *Drosophila* wing. *Curr. Biol.* **10**, 931–934 (2000).
- Cagan, R. L. & Ready, D. F. Notch is required for successive cell decisions in the developing *Drosophila* retina. *Genes Dev.* **3**, 1099–1112 (1989).
- Baker, N. E. & Zitron, A. E. *Drosophila* eye development: Notch and Delta amplify a neurogenic pattern conferred on the morphogenetic furrow by Scabrous. *Mech. Dev.* **49**, 173–189 (1995).
- Fehon, R. G. *et al.* Molecular interactions between the protein products of the neurogenic loci Notch and Delta, two EGF-homologous genes in *Drosophila*. *Cell* **61**, 523–534 (1990).
- Rabinow, L. & Birchler, J. A. Interactions of vestigial and scabrous with the Notch locus of *Drosophila melanogaster*. *Genetics* **125**, 41–50 (1990).
- Lee, E. C. & Baker, N. E. gp300sca is not a high affinity Notch ligand. *Biochem. Biophys. Res. Commun.* **225**, 720–725 (1996).
- Kidd, S., Baylies, M. K., Gasic, G. P. & Young, M. W. Structure and distribution of the Notch protein in developing *Drosophila*. *Genes Dev.* **3**, 1113–1129 (1989).
- Fehon, R. G. *et al.* Complex cellular and subcellular regulation of Notch expression during embryonic and imaginal development of *Drosophila*: Implications for Notch function. *J. Cell. Biol.* **113**, 657–669 (1991).
- Wesley, C. S. Notch and wingless regulate expression of cuticle patterning genes. *Mol. Cell Biol.* **19**, 5743–5758 (1999).
- Hart, A. C., Krämer, H. & Zipursky, S. L. Extracellular domain of the boss transmembrane ligand acts as an antagonist of the sev receptor. *Nature* **361**, 732–736 (1993).
- Parks, A. L., Klueg, K. M., Stout, J. R. & Muskavitch, M. A. Ligand endocytosis drives receptor dissociation and activation in the Notch pathway. *Development* **127**, 1373–1385 (2000).
- Lieber, T., Kidd, S., Alcamo, E., Corbin, V. & Young, M. W. Antineurogenic phenotypes induced by truncated Notch proteins indicate a role in signal transduction and may point to a novel function for Notch in nuclei. *Genes Dev.* **7**, 1949–1965 (1993).
- Rebay, L. *et al.* Specific EGF repeats of Notch mediate interactions with Delta and Serrate: implications for Notch as a multifunctional receptor. *Cell* **67**, 687–699 (1991).
- Ju, B.-G. *et al.* Fringe forms a complex with Notch. *Nature* **405**, 191–195 (2000).

27. Jennings, B., Preiss, A., Delidakis, C. & Bray, S. The Notch signalling pathway is required for *Enhancer of split* bHLH protein expression during neurogenesis in the *Drosophila* embryo. *Development* **120**, 3537–3548 (1994).
28. Wesley, C. S. & Saez, L. Analysis of Notch lacking the carboxyl terminus identified in *Drosophila* embryos. *J. Cell. Biol.* **149**, 683–696 (2000).
29. Blaumueller, C. M. *et al.* Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane. *Cell* **90**, 281–291 (1997).

Acknowledgements

We thank S. Artavanis-Tsakonas, N. Baker, K. Fischbach, T. Lieber, M. Mlodzik, M. Muskavitch and M. Young for strains and reagents; L. Saez for help with S2 cell transfections; S. Kidd for cell-surface biotinylation protocol; R. Kopan, E. Newberry, D. Towler, D. Ornitz, M. Tondravi, Y. Kasai, J. Skeath and members of the Cagan laboratory for discussions and comments on the manuscript. This work was supported by the NIH, NSF, a Research Associate Fellowship from the McDonnell Center for Cellular and Molecular Neurobiology (P.A.P.), and an NIH grant to M. Young (C.S.W.).

Correspondence and requests for materials should be addressed to R.L.C. (e-mail: cagan@molecool.wustl.edu).

Polarity controls forces governing asymmetric spindle positioning in the *Caenorhabditis elegans* embryo

Stephan W. Grill^{*†}, Pierre Gönczy^{*‡}, Ernst H. K. Stelzer^{*} & Anthony A. Hyman[†]

^{*} European Molecular Biology Laboratory (EMBL), D-69117 Heidelberg, Germany
[†] Max-Planck-Institute for Cell Biology and Genetics (MPI-CBG), D-01307 Dresden, Germany

Cell divisions that create daughter cells of different sizes are crucial for the generation of cell diversity during animal development¹. In such asymmetric divisions, the mitotic spindle must be asymmetrically positioned at the end of anaphase^{2,3}. The mechanisms by which cell polarity translates to asymmetric spindle positioning remain unclear. Here we examine the nature of the forces governing asymmetric spindle positioning in the single-cell-stage *Caenorhabditis elegans* embryo. To reveal the forces that act on each spindle pole, we removed the central spindle in living embryos either physically with an ultraviolet laser microbeam, or genetically by RNA-mediated interference of a kinesin⁴. We show that pulling forces external to the spindle act on the two spindle poles. A stronger net force acts on the posterior pole, thereby explaining the overall posterior displacement seen in wild-type embryos. We also show that the net force acting on each spindle pole is under control of the *par* genes that are required for cell polarity along the anterior–posterior embryonic axis. Finally, we discuss simple mathematical models that describe the main features of spindle pole behaviour. Our work suggests a mechanism for generating asymmetry in spindle positioning by varying the net pulling force that acts on each spindle pole, thus allowing for the generation of daughter cells with different sizes.

The first cleavage division of *C. elegans* embryos generates a large anterior and a smaller posterior blastomere along the anterior–posterior axis⁵. Polarity in this system is established after fertilization⁶ by the concerted action of at least six *par* genes⁷. During anaphase B, the spindle elongates asymmetrically: the anterior spindle pole remains in a relatively fixed position along the anterior–posterior axis, while the posterior spindle pole is

displaced towards the posterior of the embryo as it oscillates transversely⁸.

Two types of microtubule-dependent forces contribute to spindle positioning and elongation during anaphase B in other systems^{9–12}. First, overlapping spindle microtubules can generate forces that ‘push’ spindle poles apart⁹. Second, forces transmitted by astral microtubules can ‘pull’ spindle poles apart^{10–12}. To test which of these apply to the one-cell *C. elegans* embryo, we removed the spindle midzone that connects the two spindle poles at the beginning of anaphase B, and examined the resulting movement of the two independent spindle poles. If intra-spindle forces alone drive anaphase B, then the two spindle poles should not separate after the spindle is severed. In contrast, if extra-spindle pulling forces participate in anaphase B, separation of the two spindle poles should still occur after severing. The spindle midzone was severed with a pulsed ultraviolet laser microbeam, or removed by RNA mediated interference (RNAi)⁴ of a *C. elegans* kinesin that is related to XKCM1/MCAK^{13–16} (referred to as CeMCAK). In both experimental situations, we verified the disappearance of the spindle midzone by immunofluorescence with anti-tubulin antibodies (Fig. 1b, c; compare with Fig. 1a, arrows). The effect was spatially confined, as astral microtubules were not affected (Fig. 1b, c; compare with Fig. 1a, arrowheads).

We tracked each spindle pole using time-lapse differential interference contrast (DIC) microscopy. This showed both a dramatic increase in pole to pole distance (Fig. 2b, c; compare with Fig. 2a, arrowheads) and an increase in peak velocities of the spindle poles (Fig. 3) after removal of the spindle. Together, these results demonstrate that forces external to the spindle act on the spindle poles during anaphase B. Notably, in both of the experimental situations, the posterior spindle pole behaved differently compared with the anterior one (Figs 2 and 3). The posterior pole covered a greater distance, travelled at about a 40% higher peak velocity, and underwent transverse oscillations in the proximity of the cell cortex. This suggests that a greater pull acts on the posterior spindle pole than on its anterior counterpart, which may explain the overall posterior displacement of the spindle during anaphase B in wild-type embryos. Formally, a change in viscous drag could account for the observed differences in peak velocities, although this is unlikely, on the basis of an analysis of yolk-granule motion (see Methods).

We next examined how the *par* genes influence the net pulling

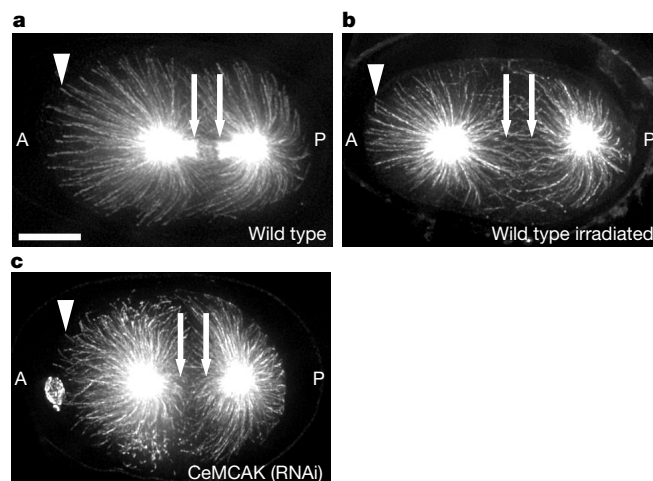


Figure 1 Spindle midzone viewed by indirect immunofluorescence with anti-tubulin antibodies. All embryos are in anaphase B. Anterior (A) is on the left and posterior (P) is on the right in this and all other figures. Scale bar, 10 μ m. **a**, In wild-type embryos, both the spindle microtubules (arrows) and astral microtubules (arrowheads) are visible. **b**, **c**, Both in the wild-type irradiated and in the CeMCAK (RNAi) embryos astral microtubules are visible (arrowheads), but spindle microtubules are not (arrows).

[‡] Present address: Swiss Institute for Experimental Cancer Research (ISREC), CH-1066 Lausanne, Switzerland.

force that acts on each spindle pole. We focused on *par-2* and *par-3*. In embryos that are mutant for either of the genes, the spindle is symmetrically positioned throughout anaphase B, presumably as a consequence of defects in establishing polarity along the anterior–posterior axis¹⁷. In wild-type embryos, PAR-3 is restricted to the anterior cortex¹⁸, and PAR-2 to the posterior cortex¹⁹. In *par-2* mutants, PAR-3 is found both at the posterior and anterior cortex of the embryo, which leads to a cortex that has anterior character throughout^{7,18,20}. Conversely, in *par-3* mutant embryos, PAR-2 fills the entire cortex of the embryo, which subsequently has posterior character throughout^{19,20}.

Our spindle-severing experiments in the wild-type embryo predict that in a *par-2* mutant embryo, the pull acting on both of the spindle poles should be equal, and resemble that exerted on the anterior spindle pole of wild-type embryos. In a *par-3* mutant embryo, the pull acting on both of the spindle poles should be equal, and resemble that exerted on the posterior spindle pole of wild-type embryos. We tested these predictions by severing the spindle midzone with an ultraviolet laser microbeam. Strikingly, in severed *par-2* mutant embryos, the resulting peak velocities of both spindle poles resembled that of the anterior spindle pole in wild-type irradiated embryos (Fig. 3). Conversely, in severed *par-3* mutant embryos, peak velocities of both spindle poles resembled that of the

posterior spindle pole in wild-type irradiated embryos (Fig. 3). Furthermore, spindle breakage events still occurred when both PAR-3 and CeMCAK were removed, whereas none were observed when the expression of both *par-2* and CeMCAK was abolished by RNAi (see Supplementary Information). This indicates that the net forces acting on both of the spindle poles in the absence of *par-2* function are not strong enough to trigger spindle rupture. All these observations show that the first division in *par-2* and *par-3* mutant embryos is symmetric because of distinct alterations in forces pulling on the spindle poles, both of which lead to equally strong pulling forces being exerted on either side.

We have shown that polarity translates into asymmetric spindle positioning by modulating the net pulling force that acts on each spindle pole; in wild-type embryos, the net pull on the posterior pole is more extensive, which leads to a displacement of the spindle towards the posterior. Force generation may be actin-dependent, although this is unlikely, as disruption of actin filaments by cytochalasin during anaphase B does not affect spindle positioning²¹. Pulling forces are more likely to be mediated by astral microtubules¹¹, perhaps by coupling force generation to microtubule depolymerization at the cell cortex^{22–24}. How the PAR proteins modulate cytoskeletal behaviour to control the net pulling force that acts on each spindle pole remains unknown. Any model

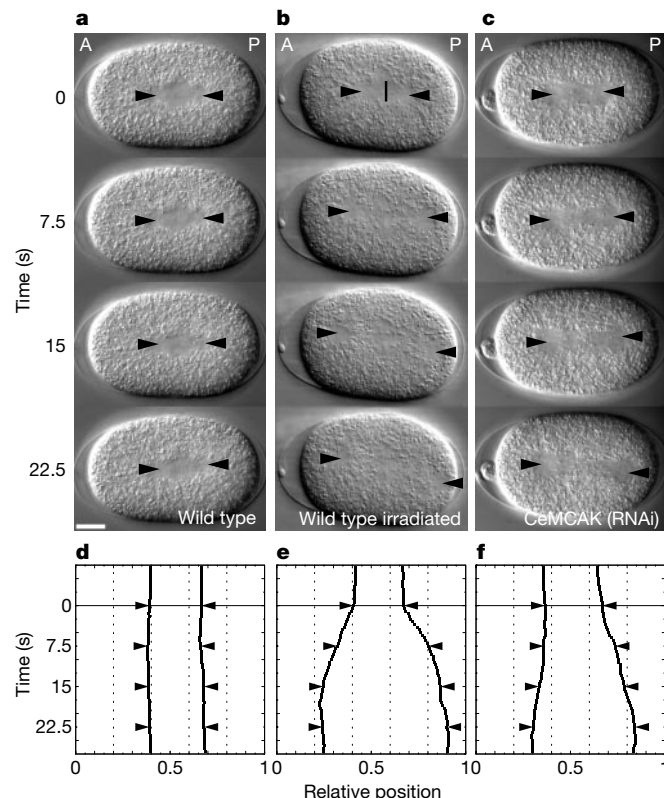


Figure 2 Pole-to-pole distance increases after spindle severing. **a–c**, Differential interference contrast (DIC) image series of *C. elegans* embryos. Spindle poles are indicated (arrowheads). Scale bar, 10 μm . **a**, Wild-type embryo. **b**, Irradiated wild-type embryo. The bar indicates where the spindle was destroyed. **c**, CeMCAK (RNAi) embryo. Spindle breakage takes place at 0 s. For both **b** and **c**, the posterior spindle pole travels further and faster than the anterior one, and a slight transverse displacement is seen in the last frame, corresponding to the beginning of transverse oscillations. Removal of ZEN-4 also results in diminished midzone microtubules without affecting anaphase B movements^{29,30}. However, the timing of spindle breakage seems to be different, leading to a more subtle effect on spindle-pole velocity during anaphase (see Supplementary Information). **d–f**, Corresponding traces of relative spindle-pole position along the anterior–posterior axis, arrowheads indicate time points for which frames in **a–c** are displayed.

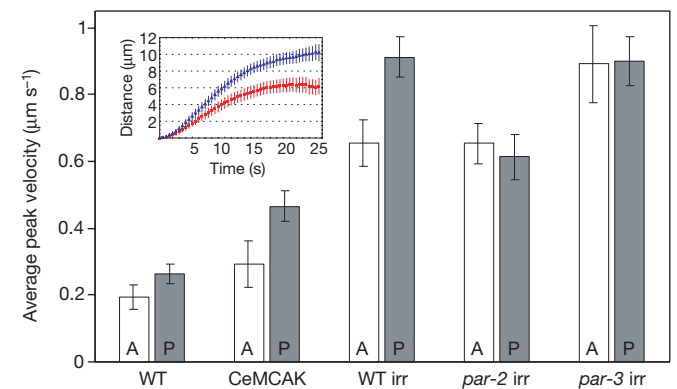


Figure 3 Average peak velocities of spindle poles increase after spindle severing. All error columns are s.e.m. with a confidence interval of 0.95. Bars show the average peak velocities of spindle poles along the anterior–posterior axis. White and black bars indicate the anterior and posterior spindle pole, respectively. Average peak velocities of spindle poles during anaphase B in wild-type (WT, $n = 20$), CeMCAK (RNAi) (CeMCAK, $n = 20$), wild-type irradiated (WT irr, $n = 34$), *par-2* (*it5*) irradiated (*par-2* irr, $n = 30$) and *par-3* (*it71*) irradiated embryos (*par-3* irr, $n = 20$) are shown. Inset, average displacement of anterior spindle pole (red) and posterior spindle pole (blue) after irradiation in wild-type embryos ($n = 34$); actual movements are in opposite directions.

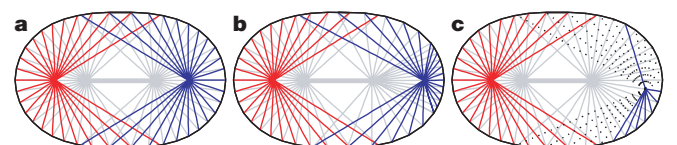


Figure 4 Numerical simulation of spindle-pole behaviour after spindle severing. The anterior aster (red) and posterior aster (blue) are shown. Grey indicates the initial state just before severing. Colours indicate the state after 22.5 s. **a**, The posterior microtubules all exert a greater pulling force than the anterior ones. **b**, All microtubules generate the same force, but the density of microtubules is increased at the very posterior end. **c**, All microtubules generate the same pulling force, but the cortical connection is weak for all microtubules of the posterior aster: they detach from the cortex instead of getting longer. Dashed lines indicate microtubules that are inactive for 12 s. Only **c** reproduces the experimental behaviour, as the posterior spindle pole travels further and faster than the anterior one, and undergoes transverse oscillations (compare with Fig. 2b and c).

describing net force generation must account for three features of spindle-pole behaviour after spindle severing (see Figs 2 and 3): the posterior spindle pole travels further than the anterior one, it travels faster, and it undergoes transverse oscillations in the proximity of the cell cortex. One possible mechanism may be that all microtubules of the posterior aster generate pulling forces that are stronger than that of the anterior aster. However, numerical simulations of this model only satisfy one of the three features, that being the posterior spindle pole travelling faster than the anterior one (Fig. 4a; see also Methods and Supplementary Information). A different hypothesis invokes an increase of forces for a subset of microtubules of the posterior aster or, similarly, an increase of microtubule density in a specific region at the posterior. Numerical simulations of this model show faster movement and a greater distance travelled by the posterior pole (Fig. 4b and Methods), but no transverse oscillations. A simpler hypothesis is that all microtubules of both asters generate equal forces, but that the interactions between microtubules and the posterior cortex are weaker than those of the anterior cortex. As an example, if microtubules of the posterior aster were to detach from the cortex instead of elongating, the net force would increase in the direction of movement. Numerical simulations of this model satisfy all three features (Fig. 4c; see also Methods).

Notably, PAR-3 has been proposed to anchor or stabilize microtubules^{18,20,25}, and it is absent from the posterior cortex, suggesting a possible mechanism for differential microtubule stability. Although the underlying molecular nature remains to be determined, our model suggests a plausible mechanism of how polarity controls cytoskeletal behaviour to ultimately generate two daughter cells of different sizes. □

Methods

Culture conditions, strains and time-lapse recordings

Basic methods of *C. elegans* culture and handling were used as described²⁶. We used strains carrying the following mutations: *daf-7* (e1372) *par-2* (it5) III (ref. 17) and *par-3* (it71) *unc-32* (e189)/qC1 (refs 20, 27). Embryos that were derived from homozygous mutant mothers are referred to as mutant embryos. We grew wild-type and *par-3* nematodes at 16 °C. *par-2* nematodes were grown at 16 °C until L3, and then shifted to 25 °C. We followed standard procedures for sample preparation and time-lapse DIC recordings at two frames per second²⁸. Polarity of the embryos was determined in all cases by examining the position of the male pronucleus, which defines the posterior of the embryo⁶.

Spindle severing

Experiments were performed using an ultraviolet laser microdissection apparatus (P.A.L.M. Mikrolaser Technologie GmbH). We focused the pulsed N₂ laser ($\lambda = 337$ nm) to a ~ 3 - μ m spot in the focal plane. *C. elegans* embryos were mounted on the inverted microscope, and as soon as a slight transverse movement of the posterior spindle pole became visible (indicating the beginning of anaphase B) about 10 shots at 30 Hz were taken at the spindle midzone. If the same number of shots were taken at regions between the spindle and the cell cortex, an increase of pole-to-pole distance was not observed (data not shown). All experimental results were verified by irradiating either the posterior or the anterior spindle pole and observing the behaviour of the other spindle pole respectively (data not shown).

RNAi of the *C. elegans* kinesin CeMCAK

We performed RNAi of the kinesin CeMCAK (gene K11D9.1) using standard procedures¹⁶, and using the primers 5'-AATTAACCTCACTAAAGGTCTGTTCGTATGGCTCCTC-3' and 5'-TAATACGACTCACTATAGGTCCTCTTTGAGCCCAAC-3'.

Single-embryo indirect immunofluorescence

Indirect immunofluorescence of single embryos was carried out with some modifications to allow for visualizing single embryos at defined stages. One pronuclear-migration-stage embryo was pipetted on a 24 $\times$ 60 mm² coverslip that was previously coated with 1% poly-L-lysine (Sigma, P1524) in PBS. The embryo was imaged on the inverted microscope, the spindle midzone was irradiated (when applicable) and the embryo then fixed by a 'freeze-crack' procedure as described²⁸. We followed standard methods²⁸ for antibody staining. Images were recorded using a widefield DeltaVision microscope (Applied Precision) and deconvolved. Figure 1 shows 2- μ m projections.

Data acquisition and analysis

The positions of both of the spindle poles were manually tracked in the DIC recordings.

We smoothed tracking data using an eight-frame running average filter, as this gave consistent results for repeated tracking of movies.

Analysis of yolk-granule motion

The observed difference in peak velocities of the anterior in comparison to the posterior spindle pole after spindle severing may be attributed to a change in viscous drag. As yolk granules probably have a large function in hindering spindle-pole movement, a mean square displacement (MSD) analysis was performed for 182 yolk granules located close to the anterior or the posterior cell cortex. Granules were automatically tracked every 0.5 s for a duration of 25 s. On a 5-s timescale, MSD plots were characteristic for either normal diffusion or for simultaneous diffusion and flow. We determined diffusion coefficients for the normal diffusing granules; their average diffusion coefficient at the anterior was less than 6% higher than their average diffusion coefficient at the posterior (data not shown). As diffusive properties of granules at the anterior versus the posterior are alike, and because spindle-pole structures are of similar size (Fig. 1a), it is unlikely that a $\sim 40\%$ difference in peak velocities only results from a difference in viscous drag.

Spindle-pole behaviour after spindle severing

In these descriptive two-dimensional models of spindle-pole behaviour, we assume that forces driving spindle-pole movement after severing are mediated by astral microtubules. We also assume that astral microtubules connect to a fixed position on the cell cortex (see Fig. 1a, arrowhead), and that they exert an equal pulling force on the spindle pole in the direction of the cortical connection. The evidence that microtubules are exerting pulling forces comes from experiments where regions between the spindle pole and the cell cortex are irradiated, which results in displacement of the spindle pole towards the opposite cell cortex (data not shown). The simulation is carried out by iteratively performing three steps. First, the net force acting on each spindle pole is calculated by summing up the contributing forces of the 25 individual microtubules. Second, the spindle pole is displaced proportionally to the net force, consistent with movement in a completely damped, viscous regime. Third (Fig. 4c only), microtubules that have increased length as a result of spindle-pole displacement are inactivated for a latency time of 12 s.

Received 18 October; accepted 24 November 2000.

- Horvitz, H. R. & Herskowitz, I. Mechanisms of asymmetric cell division: two Bs or not two Bs, that is the question. *Cell* **68**, 237–255 (1992).
- Rappaport, R. Cytokinesis in animal cells. *Int. Rev. Cytol.* **31**, 169–213 (1971).
- Gönczy, P. & Hyman, A. A. Cortical domains and the mechanisms of asymmetric cell division. *Trends Cell Biol.* **6**, 382–387 (1996).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
- Goldstein, B. & Hird, S. N. Specification of the anteroposterior axis in *Caenorhabditis elegans*. *Development* **122**, 1467–1474 (1996).
- Kemphues, K. J. & Strome, S. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 335–359 (Cold Spring Harbor Laboratory, New York, 1997).
- Albertson, D. Formation of the first cleavage spindle in nematode embryos. *Dev. Biol.* **101**, 61–72 (1984).
- Leslie, R. J. & Pickett, H. J. Ultraviolet microbeam irradiations of mitotic diatoms: investigation of spindle elongation. *J. Cell Biol.* **96**, 548–561 (1983).
- Aist, J. R. & Berns, M. W. Mechanics of chromosome separation during mitosis in *Fusarium* (Fungi imperfecti): new evidence from ultrastructural and laser microbeam experiments. *J. Cell Biol.* **91**, 446–458 (1981).
- Aist, J. R., Liang, H. & Berns, M. W. Astral and spindle forces in PtK2 cells during anaphase B: a laser microbeam study. *J. Cell Sci.* **104**, 1207–1216 (1993).
- Waters, J. C., Cole, R. W. & Rieder, C. L. The force-producing mechanism for centrosome separation during spindle formation in vertebrates is intrinsic to each aster. *J. Cell Biol.* **122**, 361–372 (1993).
- Wordeman, L. & Mitchison, T. J. Identification and partial characterization of mitotic centromere-associated kinesin, a kinesin-related protein that associates with centromeres during mitosis. *J. Cell Biol.* **128**, 95–104 (1995).
- Wein, H., Foss, M., Brady, B. & Cande, W. Z. DSK1, a novel kinesin-related protein from the diatom *Cylindrotheca fusiformis* that is involved in anaphase spindle elongation. *J. Cell Biol.* **133**, 595–604 (1996).
- Desai, A., Verma, S., Mitchison, T. J. & Walczak, C. E. Kin I kinesins are microtubule-destabilizing enzymes. *Cell* **96**, 69–78 (1999).
- Gönczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
- Kemphues, K. J., Priess, J. R., Morton, D. G. & Cheng, N. Identification of genes required for cytoplasmic localization in early *C. elegans* embryos. *Cell* **52**, 311–320 (1988).
- Etemad-Moghadam, B., Guo, S. & Kemphues, K. J. Asymmetrically distributed PAR-3 protein contributes to cell polarity and spindle alignment in early *C. elegans* embryos. *Cell* **83**, 743–752 (1995).
- Boyd, L., Guo, S., Levitan, D., Stinchcomb, D. T. & Kemphues, K. J. PAR-2 is asymmetrically distributed and promotes association of P granules and PAR-1 with the cortex in *C. elegans* embryos. *Development* **122**, 3075–3084 (1996).
- Cheng, N. N., Kirby, C. M. & Kemphues, K. J. Control of cleavage spindle orientation in *Caenorhabditis elegans*: the role of the genes *par-2* and *par-3*. *Genetics* **139**, 549–559 (1995).
- Hill, D. P. & Strome, S. An analysis of the role of microfilaments in the establishment and maintenance of asymmetry in *Caenorhabditis elegans* zygotes. *Dev. Biol.* **125**, 75–84 (1988).
- Hyman, A. A. & White, J. G. Determination of cell division axes in the early embryogenesis of *Caenorhabditis elegans*. *J. Cell Biol.* **105**, 2123–2135 (1987).

23. Lee, L. *et al.* Positioning of the mitotic spindle by a cortical-microtubule capture mechanism. *Science* **287**, 2260–2262 (2000).
24. Korinek, W. S., Copeland, M. J., Chaudhuri, A. & Chant, J. Molecular linkage underlying microtubule orientation toward cortical sites in Yeast. *Science* **287**, 2257–2259 (2000).
25. Rose, L. S. & Kempthues, K. The let-99 gene is required for proper spindle orientation during cleavage of the *C. elegans* embryo. *Development* **125**, 1337–1346 (1998).
26. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
27. Morton, D. G., Roos, J. M. & Kempthues, K. J. *par-4*, a gene required for cytoplasmic localisation and determination of specific cell types in *Caenorhabditis elegans* embryogenesis. *Genetics* **130**, 771–790 (1992).
28. Gönczy, P. *et al.* Dissection of cell division processes in the one cell stage *Caenorhabditis elegans* embryo by mutational analysis. *J. Cell Biol.* **144**, 927–946 (1999).
29. Raich, W. B., Moran, A. N., Rothman, J. H. & Hardin, J. Cytokinesis and midzone microtubule organization in *Caenorhabditis elegans* require the kinesin-like protein ZEN-4. *Mol. Biol. Cell* **9**, 2037–2049 (1998).
30. Powers, J., Bossinger, O., Rose, D., Strome, S. & Saxton, W. A nematode kinesin required for cleavage furrow advancement. *Curr. Biol.* **8**, 1133–1136 (1998).

Supplementary information (including QuickTime movies corresponding to Figs 2 and 4) is available on Nature's World-Wide Web site (<http://www.nature.com>).

Acknowledgements

We thank K. Kempthues for mutants *it71* and *it5*; K. Oegema for experimental assistance; A. Desai, E.-L. Florin and E. Hannak for discussions; R. Pepperkok, K. Schütze and R. Schütze for supplying the laser microdissection setup; and T. Bouwmeester, A. Desai, E. Hannak, E. Karsenti, F. Nédélec and K. Oegema for help with improving the manuscript.

Correspondence and requests for materials should be addressed to A.H. (e-mail: hyman@embl-heidelberg.de).

Normal human mammary epithelial cells spontaneously escape senescence and acquire genomic changes

Serguei R. Romanov*, B. Krystyna Kozakiewicz*, Charles R. Holst*, Martha R. Stampfer†, Larisa M. Haupt* & Thea D. Tlsty*

* Department of Pathology and UCSF Comprehensive Cancer Center, University of California at San Francisco, San Francisco, California 94143-0506, USA

† Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

Senescence and genomic integrity are thought to be important barriers in the development of malignant lesions¹. Human fibroblasts undergo a limited number of cell divisions before entering an irreversible arrest, called senescence². Here we show that human mammary epithelial cells (HMECs) do not conform to this paradigm of senescence. In contrast to fibroblasts, HMECs exhibit an initial growth phase that is followed by a transient growth plateau (termed selection or M0; refs 3–5), from which proliferative cells emerge to undergo further population doublings (~20–70), before entering a second growth plateau (previously termed senescence or M1; refs 4–6). We find that the first growth plateau exhibits characteristics of senescence but is not an insurmountable barrier to further growth. HMECs emerge from senescence, exhibit eroding telomeric sequences and ultimately enter telomere-based crisis to generate the types of chromosomal abnormalities seen in the earliest lesions of breast cancer. Growth past senescent barriers may be a pivotal event in the earliest steps of carcinogenesis, providing many genetic changes that predicate oncogenic evolution. The differences between epithelial cells and fibroblasts provide new insights into the mechanistic basis of neoplastic transformation.

To analyse cell-specific differences in growth and senescence, we

characterized the *in vitro* proliferation barriers in isogenic HMECs and human mammary fibroblasts (HMFs) from healthy individuals^{3,7,8}. Similar to previous studies in human skin fibroblasts¹ and HMECs^{3–5,9}, both the epithelial and fibroblast cell populations underwent a limited number of population doublings before entering a plateau (Fig. 1, HMF phase b, HMEC phase b). This plateau in fibroblasts has been variously termed the Hayflick limit², irreversible replicative senescence, and mortality stage 1 (M1)¹. The cells enlarged in size, flattened in shape, became vacuolated (Fig. 1), and expressed senescence-associated β -galactosidase¹⁰ (SA- β -gal; data not shown).

Low incorporation of bromodeoxyuridine (BrdU) and minimal presence of MCM2 protein, a DNA replication licensing factor²⁶, indicated a low proliferative index. In addition, Annexin-V staining indicated a low death index (data not shown). Further characterization showed that human foreskin fibroblasts, pre-selection HMECs and HMFs each (1) maintained genomic integrity (Fig. 2; ref. 8); (2) maintained intact cell-cycle checkpoint control (data not shown); (3) exhibited a 2N to 4N DNA content ratio of ≈ 4 at phase b (Table 1); and (4) had mean telomere restriction fragment (TRF) lengths that were similar at senescence (Fig. 3). By the morphological, behavioural and molecular criteria described above, HMFs and HMECs senesce in a manner similar to human skin fibroblasts¹. 'M0' of HMECs thus corresponds to 'M1' of fibroblasts.

The ability of HMECs and HMFs to spontaneously overcome senescence differs by several orders of magnitude. In skin fibroblasts, senescence can last for years (at least 3 yr; T.D.T., unpublished data), cells remain viable if fed routinely¹¹, and the frequency of

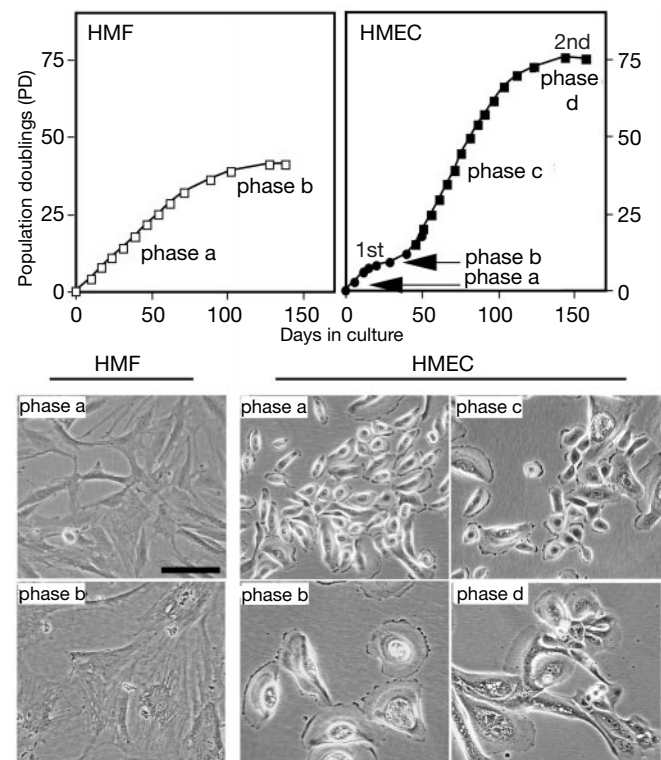


Figure 1 HMF and HMEC growth curves and cell morphologies *in vitro*. Tissue was dissociated with collagenase and hyaluronidase, and plated in parallel cultures: one in medium (DMEM) that supported the growth of fibroblasts and the other in medium (MEGM) that supported the growth of epithelial cells. The growth curve and microscopic morphology of both mammary fibroblast and epithelial cells from donor 48 during the first phase of logarithmic growth (phase a), and the first growth plateau (phase b) are shown for each population. The second epithelial phase of proliferation (phase c) and the second epithelial growth plateau (phase d) are also shown. Scale bar, 100 μ m.

spontaneous emergence is $< 10^{-9}$ (data not shown; and ref. 12). Similarly, HMFs failed to produce proliferating cells from senescent populations even after 5 months in continuous culture in either serum-containing or serum-free media ($< 6 \times 10^{-7}$; data not shown). In contrast to fibroblasts and consistent with previous reports^{3,5}, HMECs maintained at the first plateau in serum-free media sporadically emerge at a high frequency and generate clusters of small, refractile cells (Fig. 1, HMEC phase c; $1.43 \pm 0.04 \times 10^{-4}$ (donor 48, mean $\pm$ s.d., $n = 4$) and $1 \pm 1 \times 10^{-5}$ (donor 184, $n = 4$)). Pre- and post-selection HMECs showed typical heterogeneous expression of cytokeratins when examined by immuno-cytochemistry (data not shown; and ref. 13). As reported previously, HMECs that emerge from the first plateau (phase b in Fig. 1) lose expression of the p16 protein (see Supplementary Information; and refs 5, 9, 14).

After a second period of exponential growth (Fig. 1, HMEC phase c), HMECs entered a second growth plateau (Fig. 1, HMEC phase d). Unexpectedly, this plateau was critically different from the senescent, arrested state despite the fact that they both showed SA- β -gal staining (data not shown). The cells at the second plateau

displayed many hallmarks of cell crisis. HMECs at this stage were heterogeneous in size and morphology (Fig. 1, HMEC phase d). More importantly, they continued to incorporate BrdU ($16.3 \pm 1.1\%$, 4-h pulse, $n = 2$) and retained high levels of MCM2 protein ($> 50\%$ of nuclei strongly staining for MCM2; data not shown). On FACS analysis, the 2N to 4N DNA ratio was roughly 1 (data not shown), typical of a population of cells in crisis^{1,27}. This high proliferative index was counterbalanced by an increase in cell death. A significant fraction ($\sim 20\%$) of epithelial cells at the second plateau stained with Annexin-V, an indicator of cell death. In contrast, $< 1\%$ of isogenic senescent HMFs or HMECs at the first plateau were Annexin-V positive. Although DNA fragmentation characteristic of apoptosis was not detectable by TUNEL (terminal deoxynucleotidyl transferase (TdT)-mediated dUTP nick end labelling) assay, we did observe significant fragmentation of nuclei (micronucleation) in these cells as documented by DNA staining of interphase nuclei (data not shown). Thus, HMECs at the second plateau were fundamentally different from HMECs at the first plateau or fibroblast cells at senescence. They exhibited many of the characteristics of viral oncoprotein-induced crisis, with an

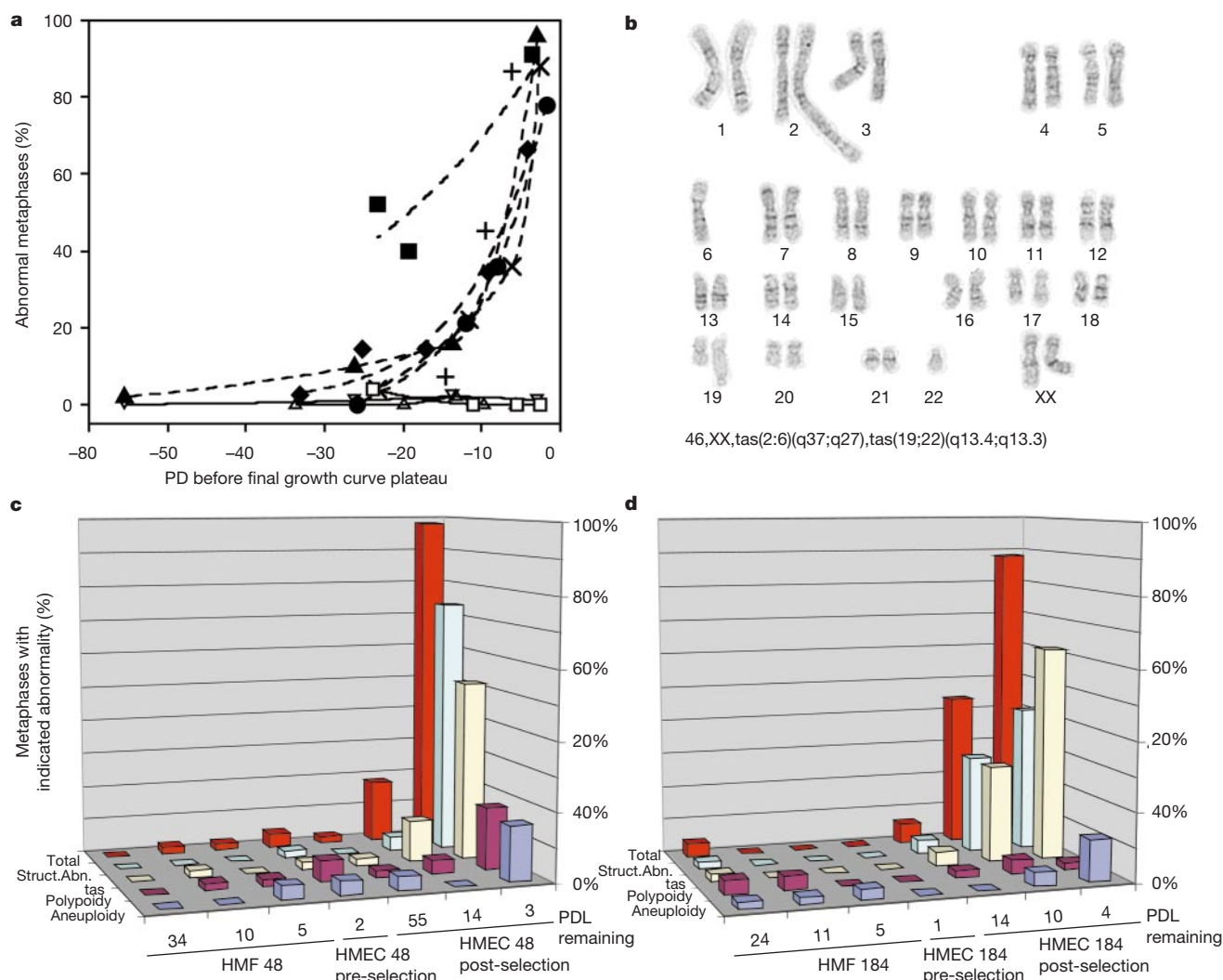


Figure 2 Spontaneous genomic instability in human mammary epithelial cells. **a**, Kinetics of accumulation of karyotypic abnormalities. The percentages of metaphase spreads with structural chromosomal abnormalities were determined as a function of the number of population doublings (PD) before final growth plateaus (0 PD). HMF populations: 48 (open triangles) and 184 (open squares). Human skin fibroblasts (open inverted triangles). HMEC populations: 48 (filled triangles), 184 spiral K (filled squares), 184 birdie (+), 4678-2 (x), 1001-3 (filled circles) and 4144-1 (filled diamonds). **b**, Representative

abnormal karyotype from HMEC 48. **c**, **d**, Types of chromosomal abnormalities observed in HMEC 48 and 184, respectively. Definitions: Total, all structural abnormalities and telomeric associations, not including numerical abnormalities; Struct. Abn., deletions, duplications, rings, marker chromosomes, chromatid exchanges and translocations; tas, telomeric associations; polyploidy, multiples of a diploid chromosome complement; aneuploidy, additions or deletions of whole chromosomes.

important exception that no immortalized variants have yet been detected.

Cytogenetic analysis of post-selection HMECs at selected passages (Fig. 1, HMEC phase c) showed that gross chromosomal abnormalities appeared in virtually every metaphase spread as the cells approached the second growth plateau (Fig. 2). In all cases, including several HMEC populations obtained commercially, the abnormalities accumulated rapidly, beginning between 10 and 20 population doublings before the final passage of cells (Fig. 2a), and coincided with slowing of the proliferation rates. In these cells, both the per cent of abnormal metaphases and the number of abnormalities per metaphase increased. The abnormalities included translocations, deletions, other rearrangements, telomeric associations, polyploidy and aneuploidy (Fig. 2b–d). Substantial polyploidy (~25%) was detected by flow analysis at final passages of post-senescent HMECs (data not shown). Microscopy revealed anaphase bridges and failed cytokineses (data not shown). The accumulation of chromosomal abnormalities was independent of donor age (range 16–50 yr) and total proliferative potential of epithelial populations (range 30–60 population doublings).

The timing and spectrum of chromosomal abnormalities, especially the numerous telomeric associations, suggested that late-passage HMECs had entered a state similar to telomere-based crisis¹⁶. Therefore, various aspects of telomere metabolism were assessed in serial subcultures of HMECs and HMFs. Both cell populations lacked telomerase activity (data not shown; and ref. 17) and exhibited a similar rate of telomere erosion (about 30 base pairs (bp) per population doubling; data not shown). As mentioned above, mean TRF lengths in isogenic HMECs at the first growth plateau and HMF at senescence were equivalent and similar to that in the earliest available passage of post-selection HMECs. Further proliferation of the post-selection HMECs was accompanied by continued shortening of the telomeres (Fig. 3a) down to a broad range of mean TRF lengths (~3.5 kb) at the second plateau. Length distribution of telomeres at each plateau was also assessed by quantitative analysis of fluorescence *in situ* hybridization (Q-FISH) of telomeric repeats¹⁵ (Fig. 3b–d). Consistent with the above determination of mean TRF lengths, mean fluorescence intensity at individual telomeres was diminished significantly (>55%) in HMECs at the second plateau as compared with those at the first plateau (Fig. 3c, d). In addition, the average number of telomeres with no signal per metaphase spread was ~4 and ~2 in senescent fibroblasts and epithelial cells at the first plateau, respectively, and increased to ~18 in epithelial cells at the second plateau. Thus, the HMECs that emerged from senescence ultimately entered a telomere-based crisis-like state.

There are striking parallels between late-generation telomerase-deficient p53 mutant mice and late-passage post-selection HMECs.

In wild-type mice, the activation of p53 mediates response to telomere dysfunction¹⁸, leading to p21 induction and cell-cycle arrest^{19,20}. Cells from the p53-mutant mice, like the post-selection HMECs, lack telomerase activity, and lack the p53-dependent arrest induced by critically shortened or dysfunctional telomeres. In both cell populations, these conditions generate similar types of chromosomal abnormalities. However, HMECs only attain this state on silencing p16 and emerging from senescence.

We assessed p53, its modulator p14^{ARF} and a downstream effector, p21, in serial subcultures of HMFs and HMECs (see Supplementary Information). Consistent with previous reports of these proteins in human skin fibroblasts^{21,22}, as HMFs grew to senescence they exhibited minimal changes in total p53 protein levels, a modest increase followed by a slight decrease of p21 protein expression, and an increase in p16 protein levels. Similar expression was seen in HMECs at the first growth plateau. c-Myc protein did not change during epithelial cell culture.

Contrary to expectations, we observed an increase in total p53 protein levels in HMECs on their emergence from senescence but not on induction of senescence (phase b) or at the second plateau (phase d). An increase in p14^{ARF} and p21, and a decrease in p16 accompanied this upregulation of p53. Both p53 and p21 proteins showed nuclear localization in post-selection cells (data not shown). Although most post-selection HMEC populations retained a constant level of p53 protein through the second plateau, only one sample (donor 48) showed further elevation of p53 protein that was not accompanied by further increases in p14^{ARF} or p21 (see Supplementary Information). These data suggest that, as HMECs emerge from senescence, unidentified signals activate the p53 pathway but fail to explain why these post-selection cells can proliferate in the presence of elevated p53 and p21.

These studies have several important ramifications. First, they challenge traditional views of how and when cells acquire genomic changes in cancer by providing a cell-intrinsic mechanism that, early in the neoplastic process, generates several simultaneous genetic changes without obligatory exposure to physical, viral or chemical mutagenic agents. Second, although post-selection HMECs have commonly been regarded as normal^{4,5}, the present observations refute this assumption. These cells do not express p16, they lack proper checkpoint control⁷ and they do not maintain genomic integrity (Fig. 2). Third, these findings redefine a high frequency spontaneous event that occurs in HMECs and show that 'M0', and not 'M1', is actually senescence. HMECs spontaneously emerge from senescence, whereas isogenic fibroblasts do not. Should these cells arise *in vivo* (a proposal consistent with preliminary observations), they would provide generative material for human carcinogenesis. Therefore, our observations that HMEC proliferation beyond senescence leads to telomeric dysfunction,

Table 1 Summary of human cell characteristics at different growth plateaux

Characteristic	Fibroblasts		Mammary epithelial cells		
	Senescence	Crisis*	1st plateau	2nd plateau	Crisis*
Lack of increase in cell number	+	+	+	+	+
SA- β -gal staining	+	+	+	+	+
BrdU incorporation	–	+	–	+	+
Cell death	–	+	–	+	+
Genomic instability	–	+	–	+	+
2N/4N ratio	>4	~1	>4	~1	?
Polyploidy	low	high	low	high	?
MCM2 expression	low	high	low	high	?
Population expansion beyond growth plateau	–	+	+	?	+
Existing nomenclature	Hayflick limit, Senescence, Irreversible replicative senescence M1	Crisis, Apparent, proliferative arrest M2	Selection, Senescence, Terminal arrest M0	Senescence, Replicative senescence M1	Crisis M2
Proposed nomenclature	Senescence	Crisis	Selection	Agonescence	Crisis

* These properties describe viral oncoprotein-induced crisis.

† See ref. 28.

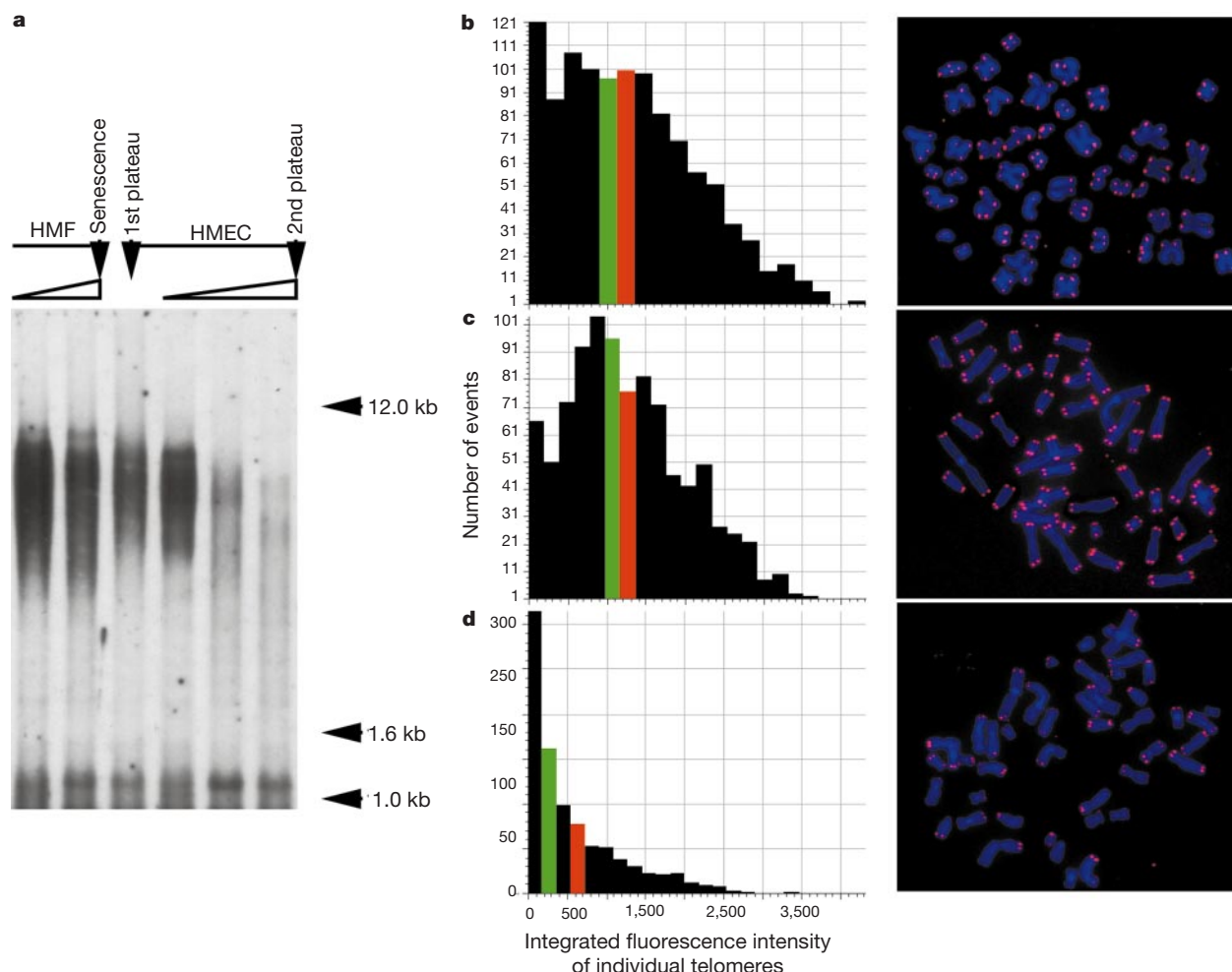


Figure 3 Post-selection HMECs continue to shorten telomeres beyond the length detected in senescent HMFs and HMECs at the first growth plateau. **a**, TRF analysis of HMF 48 (75 PD), senescent HMF 48 (42 PD), pre-selection HMEC 48 at first growth plateau (12 PD), and post-selection HMEC 48 at 20 PD, 65 PD and at second growth plateau (75 PD). **b–d**, Quantitative FISH analysis of telomeric repeats in senescent HMF 48

(**b**), HMEC 48 cells at first growth plateau (**c**) and HMEC 48 cells at second growth plateau (**d**). Shown are representative images of *in situ* hybridization of Cy3-(C₃TA)₃ PNA probe (red) to metaphase chromosomes (blue) and histograms of distribution of integrated fluorescence intensities of > 1,000 individual telomeres. Red and green bars indicate the position of mean and median values, respectively, for each data set.

coupled with observations that telomeric dysfunction leads to carcinogenesis in mice^{19,20}, might explain the early steps in carcinogenesis.

Finally, these observations identify clinical opportunities. They provide potential markers for assessing susceptibility to neoplastic transformation in individuals as well as potential targets for prevention and therapy. Several markers (Table 1; and Supplementary Information) clearly identify the different cellular states and may allow the identification of these cells *in vivo*. Remarkably, the earliest lesions in breast cancer, hyperplasias, show abnormally controlled proliferation but relatively few chromosomal structural abnormalities²³—a phenotype similar to early passage post-selection HMECs. The more progressed lesions in breast cancer, DCIS (ductal carcinoma *in situ*), also show the types of chromosomal aberrations observed in late-passage HMECs²⁴. We propose that these properties of HMECs *in vitro* are critically relevant to their transformation processes *in vivo*. Agents that minimize HMEC transition past the growth plateaux should decrease the incidence of breast cancer. Given that irreversible senescence and tight control of genomic stability are believed to be important barriers for the development of cancerous lesions, our observations also suggest that there is a much higher risk of neoplastic transformation originating in mammary epithelial tissue than in mesenchymal tissue, a prediction consistent with extensive epidemiological studies²⁵. □

Methods

Cells and cell culture

Isolation of isogenic sets of human breast epithelial cells and fibroblasts (that is, from the same gland) has been described³. Briefly, tissue from reduction mammoplasty was digested to epithelial organoids and the accompanying fibroblasts. Cells from donors 48 and 184 were obtained from a 16-year-old and a 21-year-old woman, respectively, and showed no pathologic epithelial cells. We grew and subcultured epithelial cells using two different media: MM, which contains 0.5% fetal bovine serum and several growth factors, and MEGM, a serum-free medium containing growth factors and bovine pituitary extract (BioWittaker, USA). Cell doubling times were 18–24 h in either medium. Three additional populations of human breast epithelial cells, all derived from reduction mammoplasty (1001-3, 4144-2 and 4678-2) were purchased from BioWhittaker. Fibroblasts were grown in DMEM with 10% fetal calf serum. Cells were grown at 37 °C in 5% CO₂. For routine culture, cells were counted and plated at 2×10^5 cells per 75-cm² flask. Attachment efficiency was determined by counting attached cells 15 h after plating. The number of accumulated populations doublings (PD) per passage was determined using the equation, $PD = \log[A/(BC)]/\log 2$, where *A* is the number of collected cells, *B* is the number of plated cells, and *C* is the attachment efficiency.

Chromosomal analysis

Metaphase spreads were prepared from cells treated with Colcemid (KaryoMAX, GibcoBRL, 100 ng ml⁻¹ for 6 h). We performed standard G-banding karyotypic analysis on at least 50 metaphase spreads for each population. Metaphase spreads were classified as abnormal if they contained any complement of chromosomes besides 46 XX with normal banding patterns.

Cell-cycle analysis

Cells were plated at an initial density of 5×10^4 cells per 25-cm² flask. Cells were

metabolically labelled with BrdU (10 μ M, 4 h), trypsinized, and fixed with 70% ethanol. Nuclei were isolated and stained with propidium iodide and FITC-conjugated anti-BrdU antibodies (Becton Dickinson, USA), as described⁷. Flow cytometry was performed on a FACS Sorter (Becton Dickinson). All analysed events were gated to remove debris and aggregates.

Cell death assays

TUNEL assay for DNA fragmentation was done using an In Situ Cell Death Detection kit (BMB), according to manufacturer's protocol. Alternatively, cells were stained with Annexin-V-FLUOR (BMB) and propidium iodide, and analysed by fluorescence microscopy.

Telomere length assay

Genomic DNA (10 μ g), isolated from cultured cells, was digested with restriction enzymes *RsaI* and *HinfI* and then separated in a 0.5% agarose gel. DNA was transferred to Hybond- N^+ membrane (Amersham, UK). Blots were probed with 5'-end-labelled oligonucleotide (TTAGGG)₆ and exposed to a PhosphorImager plate to detect the telomeric ends. An average telomere length was determined using ImageQuant software (Molecular Dynamics, USA).

Quantitative *in situ* hybridization with telomeric probe

In situ hybridization of telomere-specific peptide nucleic acid probe (Telomere PNA FISH Kit/Cy3, DAKO) to metaphase chromosome spreads was performed according to the manufacturer's protocol. Images were captured by a CCD camera attached to a Nikon TE300 microscope and analysed using IPLab Spectrum (Scanalytics Inc.). Background was subtracted and fluorescence signal was integrated in segments corresponding to individual telomeres¹⁵.

Senescence-associated β -galactosidase assay

Senescence-associated β -galactosidase was detected in fixed cells using a described protocol¹⁰. When staining was fully developed, the cells were washed with PBS and incubated with propidium iodide (1 μ g ml⁻¹ in PBS) and with DNase-free RNase A (5 μ g ml⁻¹). Both phase-contrast and fluorescent microscopy were performed to identify senescent cells and their nuclei.

Western analysis

Whole-cell extracts were fractionated in gradient (4–20%) polyacrylamide gels (FMC) and transferred to Hybond-P (Amersham) membrane. The following antibodies were used: mouse monoclonal anti-human p16^{ink4a} (NeoMarkers, AB-1), mouse anti-p53 (Santa Cruz, DO-1), mouse anti-p21 (WAF1) (Calbiochem, Ab-1), rabbit anti-p14^{ARF} (NeoMarkers, Ab-1) and mouse anti- α -actin (Sigma); HRP-conjugated goat-anti-mouse antibody (GibcoBRL) and goat anti-rabbit antibody (Calbiochem). Staining was developed using ECL-detection protocol.

Received 21 September; accepted 16 November 2000.

- Shay, J. W., Wright, W. E. & Werbin, H. Toward a molecular understanding of human breast cancer: a hypothesis. *Breast Cancer Res. Treatment* **25**, 83–94 (1993).
- Hayflick, L. The limited *in vitro* lifetime of human diploid cell strains. *Exp. Cell Res.* **37**, 614–636 (1965).
- Hammond, S. L., Ham, R. G. & Stampfer, M. R. Serum-free growth of human mammary epithelial cells: rapid clonal growth in defined medium and extended passage with pituitary extract. *Proc. Natl Acad. Sci. USA* **81**, 5435–5439 (1984).
- Foster, S. A. & Galloway, D. A. Human papillomavirus type 16 E7 alleviates a proliferation block in early passage human mammary epithelial cells. *Oncogene* **12**, 1773–1779 (1996).
- Huschtscha, L. I. *et al.* Loss of p16^{INK4a} expression by methylation is associated with lifespan extension of human mammary epithelial cells. *Cancer Res.* **58**, 3508–3512 (1998).
- Kiyono, T. *et al.* Both Rb/p16^{INK4a} inactivation and telomerase activity are required to immortalize human epithelial cells. *Nature* **396**, 84–88 (1998).
- Meyer, K. M., Hess, S. M., Tlsty, T. D. & Leadon, S. A. Human mammary epithelial cells exhibit a differential p53-mediated response following exposure to ionizing or UV light. *Oncogene* **18**, 5792–57805 (1999).
- Walen, K. H. & Stampfer, M. R. Chromosome analyses of human mammary epithelial cells at stages of chemical-induced transformation progression to immortality. *Cancer Genet. Cytogenet.* **37**, 249–261 (1989).
- Brenner, A. J., Stampfer, M. R. & Aldaz, C. M. Increased p16 expression with first senescence arrest in human mammary epithelial cells and extended growth capacity with p16 inactivation. *Oncogene* **17**, 199–205 (1998).
- Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).
- Pignolo, R. J., Rotenberg, M. O. & Cristofalo, V. J. Alterations in contact and density-dependent arrest state in senescent WI-38 cells. *In Vitro Cell. Dev. Biol. Anim.* **30A**, 471–476 (1994).
- Shay, J. W. & Wright, W. E. Quantitation of the frequency of immortalization of normal human diploid fibroblasts by SV40 large T-antigen. *Exp. Cell Res.* **184**, 109–118 (1989).
- Taylor-Papadimitriou, J. *et al.* Keratin expression in human mammary epithelial cells cultured from normal and malignant tissue: relation to *in vivo* phenotypes and influence of medium. *J. Cell Sci.* **94**, 403–413 (1989).
- Foster, S. A., Wong, D. J., Barrett, M. T. & Galloway, D. A. Inactivation of p16 in human mammary epithelial cells by CpG island methylation. *Mol. Cell. Biol.* **18**, 1793–1801 (1998).
- Lansdorf, P. M. *et al.* Heterogeneity in telomere length of human chromosomes. *Hum. Mol. Genet.* **5**, 685–691 (1996).
- van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).

- Stampfer, M. R. *et al.* Gradual phenotypic conversion associated with immortalization of cultured human mammary epithelial cells. *Mol. Biol. Cell* **8**, 2391–2405 (1997).
- Karlseder, J., Broccoli, D., Dai, Y., Hardy, S. & de Lange, T. p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science* **283**, 1321–1325 (1999).
- Artandi, S. E. *et al.* Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* **406**, 641–645 (2000).
- Chin, L. *et al.* p53 Deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. *Cell* **97**, 527–538 (1999).
- Alcorta, D. A. *et al.* Involvement of the cyclin-dependent kinase inhibitor p16 (INK4A) in replicative senescence of normal human fibroblasts. *Proc. Natl Acad. Sci. USA* **92**, 13742–13747 (1996).
- Hara, E. *et al.* Regulation of p16CDKN2 expression and its implications for cell immortalization and senescence. *Mol. Cell. Biol.* **16**, 859–867 (1996).
- Burbano, R. R. *et al.* Cytogenetics of epithelial hyperplasias of the human breast. *Cancer Genet. Cytogenet.* **119**, 62–66 (2000).
- Pandis, N. *et al.* Chromosome abnormalities in bilateral breast carcinomas. Cytogenetic evaluation of the clonal origin of multiple primary tumors. *Cancer* **76**, 250–258 (1995).
- Berg, J. W. & Hutter, R. V. Breast cancer. *Cancer* **75**, 257–269 (1995).
- Stoeber, K. *et al.* Cdc6 protein causes premature entry into S phase in a mammalian cell-free system. *EMBO J.* **17**, 7219–7229 (1998).
- Wei, W. & Sedivy, J. M. Differentiation between senescence (M1) and crisis (M2) in human fibroblasts cultures. *Exp. Cell Res.* **253**, 519–522 (1999).
- Tlsty, T. D. *et al.* Potentiation of genomic instability in normal human mammary epithelial cells by an epigenetic event. *J. Mammary Gland Biol. Neoplasia* (in the press).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank E. H. Blackburn, I. Herskowitz and J. Li for comments, criticism and reading the manuscript. Stimulating discussion and thoughtful critique were provided by Y. Crawford, G. Whitworth, M. Heiman, M. Springer, D. Crawford, P. Hein and J. Anderson. We thank S. Gilbert for assistance with the figures; P. Ortiz for library support; and G. Williams for MCM2 antibodies. This work was supported by NIH and NIH/NASA grants to T.D.T. and a DOE and NIH grant to M.R.S. C.R.H. is supported by a Howard Hughes Pre-doctoral Fellowship.

Correspondence and requests for materials should be addressed to T.D.T. (e-mail: ttlsty@itsa.ucsf.edu).

The bacterial conjugation protein TrwB resembles ring helicases and F₁-ATPase

F. Xavier Gomis-Rüth*, Gabriel Moncalián†, Rosa Pérez-Luque*, Ana González‡§, Elena Cabezón†, Fernando de la Cruz† & Miquel Coll*

* Institut de Biologia Molecular de Barcelona, CSIC, Jordi Girona, 18-26, 08034 Barcelona, Spain

† Departamento de Biología Molecular (Laboratorio asociado al CIB, CSIC), Universidad de Cantabria, Herrera Oria, 39011 Santander, Spain

‡ EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany

The transfer of DNA across membranes and between cells is a central biological process; however, its molecular mechanism remains unknown. In prokaryotes, *trans*-membrane passage by bacterial conjugation, is the main route for horizontal gene transfer. It is the means for rapid acquisition of new genetic information, including antibiotic resistance by pathogens. *Trans*-kingdom gene transfer from bacteria to plants¹ or fungi² and even bacterial sporulation³ are special cases of conjugation. An integral membrane DNA-binding protein, called TrwB in the *Escherichia coli* R388 conjugative system, is essential for the conjugation process. This large multimeric protein is responsible for recruiting the relaxosome DNA–protein complex, and participates in the

§ Present address: Stanford Linear Accelerator Center, 2575 Sand Hill Road, Mail Stop 99, Menlo Park, California 94025, USA.

transfer of a single DNA strand during cell mating. Here we report the three-dimensional structure of a soluble variant of TrwB. The molecule consists of two domains: a nucleotide-binding domain of α/β topology, reminiscent of RecA and DNA ring helicases, and an all- α domain. Six equivalent protein monomers associate to form an almost spherical quaternary structure that is strikingly similar to F_1 -ATPase. A central channel, 20 Å in width, traverses the hexamer.

Bacterial conjugation is mediated by proteins encoded by two gene regions of conjugative plasmids in Gram-negative bacteria. The first region encodes the DNA transfer and replication proteins (*Dtr*) and the second one encodes the mating pair formation proteins (*Mpf*). R388 is a 33-kilobase enterobacterial conjugative plasmid that contains genes that code for sulphonamide and trimethoprim resistance. It displays the simplest known *Dtr* region⁴ encoding only three proteins. TrwC acts as a relaxase as well as a DNA helicase. TrwA is a transcriptional repressor. These

two proteins comprise, together with the origin of transfer (*oriT* DNA) and the host-encoded integration host factor, the relaxosome. The third protein encoded by the R388 *Dtr* region, TrwB, is a basic (isoelectric point (pI) 10) integral membrane protein of 507 residues⁵ with a nucleotide-binding site (NBS). This feature is essential for conjugation, because a single point mutation affecting the Walker A motif (K136T) is completely deficient in gene transfer⁵.

TrwB binds single-stranded and double-stranded DNA nonspecifically and independently of NTP binding. The protein belongs to the family of TraG-like coupling proteins responsible for recruiting the relaxosome and coupling it with the structures involved in the cell-to-cell contact^{6,7}. This family displays similarity with the SpoIIIE/FtsK protein family⁵. SpoIIIE is an essential sporulation membrane-bound protein in *Bacillus subtilis* that is responsible for DNA translocation from the mother cell to the prespore³, whereas FtsK is involved in the resolution of chromosomal dimers during

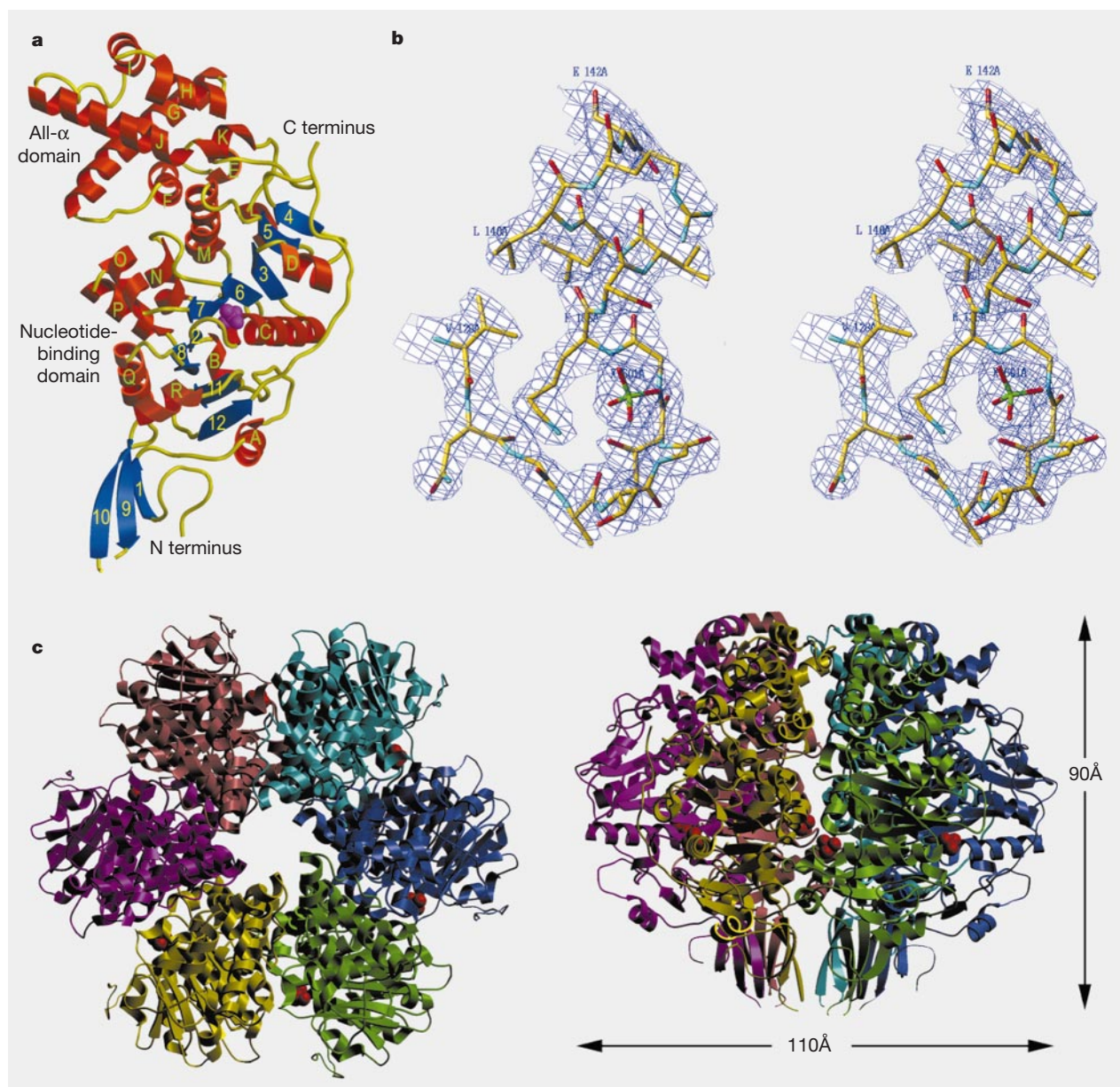


Figure 1 Structural features of TrwB. **a**, Ribbon diagram of a TrwB Δ N70 protomer. Strands are numbered (1–12), helices are labelled with capital letters (A–R). **b**, Stereo representation of the final ($2F_{\text{obs}} - F_{\text{calc}}$) map of the P3₂₁ crystal at the nucleotide-

binding site with the final model coordinates superimposed. A sulphate anion is visible at the position of a β -phosphate group of the substrate. **c**, Hexamer viewed from the cytoplasmic side. Left, view along the local six-fold axis; right, side view.

bacterial cell division⁸. SpoIIIE and FtsK are believed to function as motors that drive DNA across the membrane annulus during sporulation and cell division, respectively. In addition, TraG-like proteins are also related to VirD4, a protein from the plant tumour-inducing bacterial pathogen *Agrobacterium tumefaciens*, which is needed to deliver transforming DNA into the plant cell.

We have solved the structure of a soluble fragment of TrwB lacking the transmembrane part (TrwB Δ N70) in two crystal forms. The TrwB Δ N70 monomer displays an 'orange-segment' shape of approximate dimensions 90 $\times$ 45 $\times$ 40 Å (Figs 1a, 2a). It can be subdivided into two domains: an all- α helical domain (AAD); and an α/β twisted open-sheet domain containing the nucleotide-binding domain (NBD), which is proximal to the membrane. The NBD runs from the amino terminus, which is the anchor point of the excised 70-residue transmembrane domain, to residue Lys 183 and continues from position Asp 298 to the carboxy terminus located at the surface. The NBD is composed of a central, highly twisted (about 90°), nine-stranded β -pleated sheet of mixed parallel/antiparallel topology (strands 2–8, 11 and 12; see Fig. 1a), which is flanked on its concave side by four helices (A–D) and on its convex side by seven helices (L–R). The convex side faces the interior channel of the particle. On the membrane-proximal edge of this central β -sheet, a small three-stranded antiparallel sheet (strands 1, 9 and 10) is inserted, almost perpendicular to the former. Strands 9 and 10 are only partially defined by electron density (see Methods), probably owing to the absence of interactions with the (excised) transmembrane domain preceding strand 1. On top of the NBD, the smaller AAD is inserted between strand β 4 and helix α L, comprising residues Gly 184 to Gly 297 (Fig. 1a). This domain contains seven helices: a 3_{10} -helix (helix E), followed by a two-helix hairpin (helices F and G), followed at the C-terminal end by a

second, curved, helical hairpin segment (helices H–K), which mainly protects and interacts with the central helix G.

In both crystal forms, six TrwB Δ N70 protomers oligomerize using a local six-fold axis (Fig. 1c). This agrees both with biochemical data showing that the native protein migrates as a hexamer in gel-filtration column chromatography, and with electron microscopy images (data not shown). The hexamer has overall dimensions of 110 Å in diameter and 90 Å in height (Fig. 1c), and a roughly spherical shape, which is slightly flattened at both poles. A central channel runs from the cytosol pole (made up by the AADs) to the membrane pole (formed by the NBDs; Fig. 2a, b). The channel entrance is plugged by a ring of asparagine residues (Asn 271 of loop α I– α J from each subunit) and restricted to $\sim$ 8 Å in diameter. This is the narrowest part of the channel—which is $\sim$ 20 Å wide along its whole length and ends at its membrane side with an opening of $\sim$ 22 Å—although a narrower section may occur at the transmembrane part to avoid membrane permeability.

The interaction between TrwB Δ N70 monomers is mainly hydrophilic (Fig. 2a), and buries a common surface of 4,588 Å² (about 25% of the total surface of a monomer) with the central twisted β -pleated sheets arranged almost in parallel (C-edge to N-edge; Fig. 1c). The total intermolecular contact area of a monomer (for example, A with both B and G), is 9,100 Å², which is almost 50% of the total surface. A notable interaction is observed between strands 9 and 10 (partially interrupted at their ends) and residues of the N-terminal segments of each monomer to form a 12-stranded β -barrel. This barrel might coat the interior of the transmembrane channel formed by the helical segments of the transmembrane domain (Fig. 2c).

The NBS is located at the C-terminal edge of the sheet and is shaped by loops β 2– α C and β 6– α N. These two segments contain

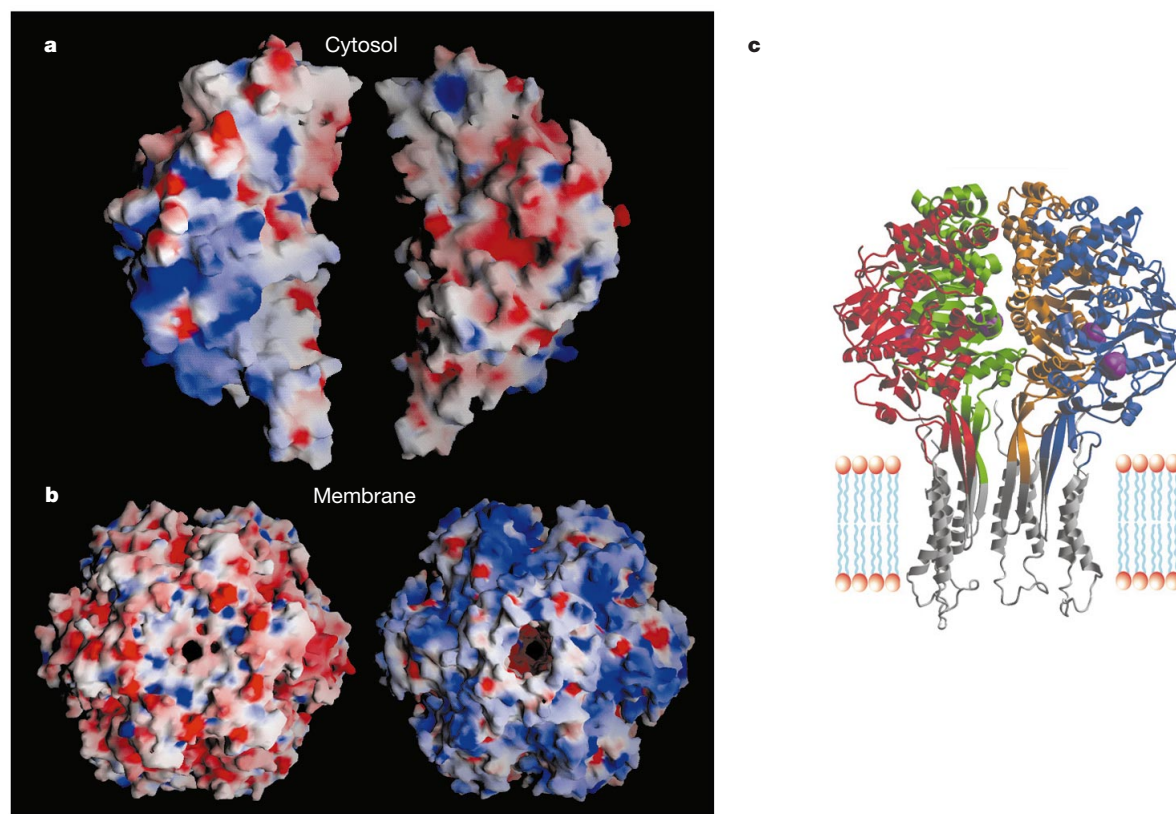


Figure 2 Electrostatic surface representations. **a**, Lateral view. Only two (opposed) protomers of the hexamer are shown to highlight the surface complementarity and the central channel. **b**, View along the six-fold axis from the cytosolic pole (left) and from the membrane pole (right). **c**, Ribbon representation of the complete TrwB, including a model

for the transmembrane part (in grey), based on two consecutive transmembrane helices of the photosynthetic reaction centre structure (PDB access code 1mps). Only four of the six protomers are shown for clarity.

the Walker A motif or P-loop (G130-A131-T132-G133-T134-G135-K136-S137) and the Walker B motif (D356-E357-L358-A359) found in NTP-binding proteins⁹. The six nucleotide-binding sites are located on superficial cavities, about 32 Å apart from each other at interfaces between vicinal protomers. We superimposed the structure of F₁-ATPase¹⁰, with which TrwB has strong structural similarity, to deduce the residues involved in nucleotide binding. Possible residues are Thr 132, whose O_γ atom would bind the γ-phosphate oxygens; Lys 136, which binds the β-phosphate oxygens; and Ser 137, which would interact through a Mg²⁺ cation with oxygens of both the β- and γ-phosphates. The proposed catalytic residues in F₁-ATPase do not match any residues in the present structure, suggesting that it has a different enzymatic mechanism. The C terminus of TraG-like proteins has been implicated in the efficiency and specificity of their interaction with the relaxosome³¹. It might be significant that the TrwB C-terminal part is directly linked to the NBS, and thus conceivably may modulate NTP hydrolysis through interaction with components of the relaxosome.

Co-crystallization and soaking trials of trigonal or monoclinic crystals in solutions containing standard NTP analogues always resulted in partially occupied NBSs, making the precise definition of inhibitor binding impossible, in particular of the base moiety. It might be that the active sites function in a sequential manner, as has been suggested for bacteriophage T7 gene 4 helicase¹¹, and that a particle with all six sites occupied is not viable. This could lead to disordered low-occupancy sites in the complexes. In the trigonal crystals, however, a strongly bound sulphate anion (100% occupancy) was found in the nucleotide-binding pocket, at the position where the nucleotide β-phosphate should be located (Fig. 1b). When we superimposed a protomer from this trigonal crystal with one of a monoclinic crystal grown from tartrate (root mean square deviation (r.m.s.d.) 0.47 Å) that has empty NBSs, we found that some regions deviate strongly, despite the close overall similarity of the hexamers. In particular, segments Gly 130 to Arg 141 (r.m.s.d. 1.1 Å) encompassing the P-loop, Gly 384 to Gln 390 (r.m.s.d. 1.0 Å), and Gly 414 to Gly 415 (r.m.s.d. 1.6 Å) diverge greatly, including their main chain atoms. Notably, the main chain of Walker motif B is not affected. These differing regions are not involved in any surface crystal contact, and should therefore be considered to be a consequence of the occupation of the NBSs. An internal salt bridge made up by Glu 357 (in Walker B motif) with the crucial Lys 136 (in Walker A motif) in the unliganded structure is disrupted in the SO₄²⁻ complex. The latter side chain is rotated to approach the substrate β-phosphate position. This is accompanied by a series of rearrangements transmitting the movement to the interior channel at position Ser 289, which is located at the beginning of helix O.

Topological searches of TrwB against proteins of known tertiary structure reveal, despite negligible sequence similarity, a high structural similarity of its NBD with the equivalent domain of RecA¹² (PDB access code 2reb) and other RecA-like core encompassing enzymes. Among these are bacteriophage T7 gene 4 helicase^{11,13} (PDB access codes 1cr0, 1eok and 1eoj), the δ'-subunit

of the clamp loader complex of *E. coli* DNA polymerase III (ref. 14) (1a5t), and the *N*-ethylmaleimide-sensitive fusion protein (NFS), a cytosolic ATPase required for intracellular vesicle fusion reactions¹⁵ (1d2n). Also, there is strong structural similarity with both α- and β-subunits of F₁-ATPase, part of the membrane-associated F₀F₁-ATPase complex responsible for energy conversion through axial rotation movement in mitochondria, chloroplasts and bacteria¹⁰ (1bmf). This structural analogy is much higher with the β-subunits in the occupied conformation (β_{DP} and β_{TP}) than with the empty one (β_E).

When comparing TrwB with all the hexameric ring quaternary structures, helicases and NFS appear to be slightly wider and more flat-topped, although this effect is overemphasized because of missing domains in the solved structures. TrwB is almost spherical and remarkably similar, in shape and dimensions, to the F₁-ATPase α₃β₃ heterohexamer, with which it shares the property of being a membrane-associated protein. In contrast, TrwB is a homohexamer. Several representatives of the helicase/ATPase-like proteins display an additional AAD, besides the core NBD, as shown for RecA, NSF, *Bacillus stearothermophilus* DExx box DNA helicase¹⁶, δ'-subunit of the clamp loader complex of *E. coli* DNA polymerase III and α- and β-subunits of F₁-ATPase, among others. These are arranged in different ways with respect to the NBD; however, detailed topological inspections reveal that TrwB AAD bears significant structural similarity only with N-terminal domain 1 of the site-specific recombinase, XerD, of the λ-integrase family (PDB access code 1a0p; ref. 17). In XerD, this all-α domain 1 has been proposed to contribute partially to DNA binding, which is achieved mainly by domain 2. Proteins in the TrwB superfamily share the NBD but have non-homologous AADs. This argues in favour of these domains being added later on in the evolution of this protein family to modulate ATP hydrolysis as a result of AAD-specific interactions.

DNA helicases use the energy from NTP hydrolysis to unwind double-stranded DNA. Those that act at the replication fork are ring-shaped hexameric enzymes that move along one DNA strand that passes through the central hole, and that catalyse the displacement of the complementary strand^{13,18–20}. The crystal structures of the two ring helicases T7 gene 4 (ref. 11) and RepA (see ref. 21) led the former authors to propose a sequential NTP hydrolytic mechanism. The strong structural resemblance of TrwB with DNA ring helicases suggests, by analogy, that the single DNA T-strand might pass through the central channel of the particle, thus being introduced into the mating/translocation apparatus that connects donor and recipient cells. ATP hydrolysis would provide the energy to pump the single-stranded DNA, much in the same way as it does in helicases for their processive movement along the DNA^{22,23}. The structure of TrwB with a sulphate ion bound at the NBS shows that this binding results in structural changes that are transmitted to the central channel. This gives an idea on how the protein might function: ATP binding/hydrolysis could trigger a molecular switch mechanism affecting the channel, and therefore the DNA binding and displacement through it. One obstacle to the hypothesis that the single-stranded DNA traverses the central channel is the asparagine plug at the cytoplasmic side, which narrows the channel to ~8 Å (Fig. 2a); elsewhere the channel is an appropriate diameter (~20 Å) to accommodate a single DNA strand. If the channel hypothesis is correct, our structure would represent a closed conformation and the cytoplasmic gate would open on activation—a movement that might be triggered by attachment of relaxosome components. □

Methods

Crystallization and heavy-atom derivation

The soluble 437-residue variant of TrwB (sequence accession number SWALL:Q04230), TrwBΔN70, was prepared as described². The protein was crystallized from drops initially containing 1.5 M ammonium sulphate, 0.1 M NaCl, 0.1 M HEPES pH 7.5. Trigonal crystals belonging to space group P3₁21 (*a* = *b* = 151.3 Å, *c* = 258.2 Å) were obtained. These crystals diffract to about 2.0 Å resolution and contain six protomers per asymmetric unit.

Table 1 Final refinement and model statistics

Structure	SO ₄ ²⁻ complex	Apo
Space group	P3 ₁ 21	P2 ₁
Resolution range (Å)	50–2.4	50–2.5
No. of reflections used	133,805	176,026
Crystallographic <i>R</i> _{factor} (free <i>R</i> _{factor}) [*]	21.0 (24.8)	20.9 (26.7)
No. of protein atoms (a.u.)	19,693	41,537
No. of solvent molecules	965	1,559
No. of sulphate anions	15	–
R.m.s. deviation from target values		
Bonds (Å)	0.008	0.008
Angles (°)	1.393	1.387
Bonded <i>B</i> factors (Å ²)	1.826	1.959
Average <i>B</i> factors for protein atoms (Å ²)	43.0	40.8

^{*} *R*_{factor} = Σ_{hkl} ||*F*_{obs}|| – *k*||*F*_{calc}||/Σ_{hkl} ||*F*_{obs}||; free *R*_{factor}, same for a test set of 7% reflections (>500) not used during refinement (until last-but-one cycle).

When 0.9 M potassium/sodium tartrate, 0.1 M HEPES pH 7.5 was used as precipitant, similar trigonal crystals appeared in the drops together with a second, monoclinic ($P2_1$) crystal form ($a = 107.4 \text{ \AA}$, $b = 153.4 \text{ \AA}$, $c = 162.5 \text{ \AA}$, $\beta = 94.2^\circ$). These crystals diffract as well as the trigonal although they contain 12 protomers in the asymmetric unit. A tantalum bromide ($\text{Ta}_6\text{Br}_{12}^{2+}$) derivative of the trigonal crystal form was obtained by overnight soaking of native crystals resulting in a unit cell with an enlarged c axis parameter ($a = b = 151.2 \text{ \AA}$, $c = 262.8 \text{ \AA}$).

Data collection and phasing

Complete diffraction datasets were collected from a single crystal each at EMBL beam lines BW7A and BW7B (Outstation Hamburg, DESY) and ID14-2 and ID14-4 (Outstation Grenoble, ESRF). Three datasets were recorded from the $\text{Ta}_6\text{Br}_{12}^{2+}$ derivative (peak, inflection point and hard remote) around the theoretical Ta L-III edge. All data were processed with MOSFLM 6.01 (ref. 24) and scaled, merged and reduced with programs of the CCP4 suite²⁵. A self-rotation calculation performed for the trigonal data indicated the presence of a local six-fold axis at Eulerian angles $\alpha = 12.8^\circ$, $\beta = 56.6^\circ$ and $\gamma = 12.8^\circ$, in accordance with the presence of six monomers displaying local C_6 symmetry. The structure was solved by multiple-wavelength anomalous diffraction (MAD) using the tantalum bromide derivative of the trigonal crystal form using SHELXS²⁶ and MLPHARE²⁵. Six sites were found and refined, and phases were computed to $4.5 \text{ \AA}$. A subsequent density modification step using DM within CCP4 (ref. 25) rendered an F_{obs} electron density map that allowed us to build a monomer mask and localize the local six-fold axis which was consistent with the self-rotation function. The original MAD-phases to $4.5 \text{ \AA}$ were used with the native structure factor amplitudes and a density modification run with non-crystallographic-symmetry (NCS) averaging and phase extension to $2.7 \text{ \AA}$ was computed. This led to a straightforward traceable native map. The model was built with Turbo-Frodo.

Structure refinement

Successive cycles of maximum-likelihood positional and temperature factor refinement with CNS version 1.0 (ref. 27) using progressively all data up to the full resolution of $2.4 \text{ \AA}$ and keeping NCS restraints followed. Computation of phased-combined maps and manual model building permitted the gradual completion of the model. At the final stages, solvent molecules and sulphate ions were introduced. Table 1 provides a summary for final model refinement. The structure of the monoclinic crystal form was solved by molecular replacement with AMoRe²⁸. A whole TrwB Δ N70 hexamer was used as a searching model and two clear solutions were obtained with correlation coefficient and R_{factor} equal to $67.4/36.3\%$ (second highest solution $41.2/47.1\%$). The model was inspected and the new electron-density-based differences with the trigonal structure were corrected. Solvent molecule position assignment and model refinement with CNS proceeded similarly and applying NCS restraints. No ions were localized in these crystals. All residues excepting His 125 of each chain, clearly defined by electron density and involved in β -sheet interactions, are in allowed regions of the Ramachandran plot. Superimpositions were calculated with Turbo-Frodo, Lsqkab of the CCP4 suite²⁵ and Lsqman of the RAVE package (Uppsala Software Factory; <http://alpha2.bmc.uu.se/~gerard/manuals>). Figures were computed with Turbo-Frodo, SETOR²⁹ and GRASP³⁰. Lee & Richards buried surface accessibility calculations (probe radius $1.4 \text{ \AA}$) and close contacts ($< 4 \text{ \AA}$) were ascertained with CNS version 1.0. Structural similarity searches were performed with the DALI server (<http://www.ebi.ac.uk/dali>).

Received 21 September; accepted 8 November 2000.

1. Stachel, S. E. & Zambryski, P. *Agrobacterium tumefaciens* and the susceptible plant cell: a novel adaptation of extracellular recognition and DNA conjugation. *Cell* **47**, 155–157 (1986).
2. Heinemann, J. A. & Sprague, G. F. J. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**, 205–209 (1989).
3. Wu, L. J., Lewis, P. J., Allmansberger, R., Hauser, P. M. & Errington, J. A conjugation-like mechanism for prespore chromosome partitioning during sporulation in *Bacillus subtilis*. *Gene Dev.* **9**, 1316–1326 (1995).
4. Llosa, M., Bolland, S. & de la Cruz, F. Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. *J. Mol. Biol.* **235**, 448–464 (1994).
5. Moncalián, G. et al. Characterization of ATP and DNA binding activities of TrwB, the coupling protein essential in plasmid R388 conjugation. *J. Biol. Chem.* **274**, 36117–36124 (1999).
6. Zechner, E. L. in *The Horizontal Gene Pool: Bacterial Plasmids and Gene Spread* (ed. Thomas, C. M.) 87–173 (Harwood Academic, London, 2000).
7. Cabezon, E., Sastre, J. I. & de la Cruz, F. Genetic evidence of a coupling role for the TraG protein family in bacterial conjugation. *Mol. Gen. Genet.* **254**, 400–406 (1997).
8. Beggs, K. J., Dewar, S. J. & Donachie, W. D. A new *Escherichia coli* cell division gene, *ftsK*. *J. Bacteriol.* **177**, 6211–6222 (1995).
9. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the α - and β -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951 (1982).
10. Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. Structure at $2.8 \text{ \AA}$ resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**, 621–628 (1994).
11. Singleton, M. R., Sawaya, M. R., Ellenberger, T. & Wigley, D. B. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**, 589–600 (2000).
12. Story, R. M., Weber, I. T. & Steitz, T. A. The structure of the *E. coli* recA protein monomer and polymer. *Nature* **355**, 318–325 (1992).
13. Sawaya, M. R., Guo, S., Tabor, S., Richardson, C. C. & Ellenberger, T. Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell* **99**, 167–177 (1999).
14. Guenther, B., Onrust, R., Sali, A., O'Donnell, M. & Kuriyan, J. Crystal structure of the δ' subunit of the clamp-loader complex of *E. coli* DNA polymerase III. *Cell* **91**, 335–345 (1997).

15. Lenzen, C. U., Steinmann, D., Whiteheart, S. W. & Weis, W. I. Crystal structure of the hexamerization domain of N-ethylmaleimide-sensitive fusion protein. *Cell* **94**, 525–535 (1998).
16. Subramanya, H. S., Bird, L. E., Brannigan, J. A. & Wigley, D. B. Crystal structure of a DExx box DNA helicase. *Nature* **384**, 379–383 (1996).
17. Subramanya, H. S. et al. Crystal structure of the site-specific recombinase, XerD. *EMBO J.* **16**, 5178–5187 (1997).
18. Egelman, E. H., Yu, X., Wild, R., Hingorani, M. M. & Patel, S. M. Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases. *Proc. Natl Acad. Sci. USA* **92**, 3869–3873 (1995).
19. Hacker, K. J. & Johnson, K. A. A hexameric helicase encircles one DNA strand and excludes the other during DNA unwinding. *Biochemistry* **36**, 14080–14087 (1997).
20. Yu, X., Hingorani, M. M., Patel, S. S. & Egelman, E. H. DNA is bound within the central hole to one or two of the six subunits of the T7 DNA helicase. *Nature Struct. Biol.* **3**, 740–743 (1996).
21. Soultanas, P. & Wigley, D. B. DNA helicases: 'inching forward'. *Curr. Opin. Struct. Biol.* **10**, 124–128 (2000).
22. Raney, K. D. & Benkovic, S. J. Bacteriophage T4 DDA helicase translocates in an unidirectional fashion on single-stranded DNA. *J. Biol. Chem.* **270**, 22236–22242 (1995).
23. Kaplan, D. L. The 3'-tail of a forked-duplex sterically determines whether one or two DNA strands pass through the central channel of a replication-fork helicase. *J. Mol. Biol.* **301**, 285–299 (2000).
24. Leslie, A. G. W. in *Crystallographic computing V* (eds Moras, D., Podjarny, A. D. & Thierry, J. C.) 27–38 (Oxford Univ. Press, Oxford, 1991).
25. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
26. Sheldrick, G. M. Patterson superposition and *ab initio* phasing. *Methods Enzymol.* **276**, 628–641 (1997).
27. Brünger, A. T. et al. Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
28. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
29. Evans, S. V. SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134–138 (1993).
30. Nicholls, A., Bharadwaj, R. & Honig, B. GRASP: graphical representation and analysis of surface properties. *Biophys. J.* **64**, A166–A166 (1993).
31. Sastre, J. I., Cabezon, E. & de la Cruz, F. The carboxyl terminus of protein TraD adds specificity and efficiency to F-plasmid conjugative transfer. *J. Bacteriol.* **180**, 6039–6042 (1998).

Acknowledgements

We are most grateful to R. Huber for making tantalum bromide available to us, and to I. Usón for help with SHELX. This work was supported by grants from the Ministerio de Educación y Cultura of Spain, the Generalitat de Catalunya and the European Union. Synchrotron data collection was supported by EU grants and the ESRF.

Correspondence and requests for materials should be addressed to M.C. (e-mail: mcccr@ibmb.csic.es). The co-ordinates of TrwB Δ N70 have been deposited with the Protein Data Bank (access codes 1e9r and 1e9s).

Three key residues form a critical contact network in a protein folding transition state

Michele Vendruscolo*, Emanuele Paci*†, Christopher M. Dobson* & Martin Karplus*‡

* Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, UK

† Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 4 rue Blaise Pascal, 67000 Strasbourg, France

‡ Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

Determining how a protein folds is a central problem in structural biology. The rate of folding of many proteins is determined by the transition state, so that a knowledge of its structure is essential for understanding the protein folding reaction. Here we use mutation measurements—which determine the role of individual residues in stabilizing the transition state^{1,2}—as restraints in a Monte Carlo sampling procedure to determine the ensemble of structures that make up the transition state. We apply this approach to the experimental data for the 98-residue protein acylphosphatase³, and obtain a transition-state ensemble with the

native-state topology and an average root-mean-square deviation of 6 Å from the native structure. Although about 20 residues with small positional fluctuations form the structural core of this transition state, the native-like contact network of only three of these residues is sufficient to determine the overall fold of the protein. This result reveals how a nucleation mechanism involving a small number of key residues can lead to folding of a polypeptide chain to its unique native-state structure.

As the transition state corresponds to the least stable region on a potential energy surface, it is difficult to determine its structure by the methods commonly used for the stable portions of the surface (that is, reactants, products or intermediates). For certain small-molecule reactions direct observation of the transition state has

been possible⁴, but it is not evident how to achieve a corresponding result for protein folding. The protein engineering method, introduced by Fersht, Winter and co-workers in 1982 (refs 5, 6), can be used to provide experimental information about the residue interactions present in the transition state for folding^{1,2}. The essential element of the method is measurement of the quantity ϕ_i^{exp} , which is the ratio of the change in stability, $\Delta\Delta G_{\text{TSE}}$, of the transition-state ensemble (TSE) to that of the native state, $\Delta\Delta G_{\text{NSE}}$, due to the mutation of residue i ; the unfolded state is used as the reference.

Here we introduce a strategy that substantially extends the information that can be derived from a set of ϕ_i^{exp} values for a protein. The essence of the method is that the TSE is restrained to be the most stable region of an energy function, which includes the information from the ϕ_i^{exp} values, rather than being an unstable region, as for the true energy function of the protein. In this way the TSE can be obtained by standard methods for sampling conformational space, in a manner analogous to the determination of native structures from experimental NMR data⁷.

Our approach uses the ϕ_i^{exp} values, interpreted in terms of native-like inter-residue contacts, as restraints to generate an ensemble of structures. It is applied here to determine the TSE of acylphosphatase (AcP) (Fig. 1a). A representation of the TSE of AcP consisting of 1,144 structures was generated using only the 24 measured ϕ_i^{exp} values³ as restraints (the 'primitive' model, see Methods). The Pearson linear correlation coefficient ρ between the experimental values and those obtained from the TSE is 0.99. (See also Supplementary Information). As the chain connectivity requires that certain contacts are formed simultaneously, the very high correlation demonstrates the internal consistency of the experimental data. Moreover, this finding provides additional evidence for the fundamental assumptions on which the ϕ -value analysis is based².

Figure 2a shows the calculated values (ϕ_i^{calc}) for all the residues and compares them with the ϕ_i^{exp} values. For most of the residues for which experimental data are not available, the ϕ_i^{calc} interpolate smoothly between the ϕ_i^{exp} . There are, however, some exceptions. The most important of these are in the central region of strand β_4 (residues 79–82), for which the predicted values are relatively large. Although the hydrophobic mutation investigated in the region of AcP (F80L) destabilized the protein too much for ϕ_i^{exp} to be determined³, ϕ values for the corresponding region of the procarboxypeptidase A2 activation domain (ADA2h), (ref. 8), a protein having the same overall topology as AcP, have been measured; the average ϕ_i^{exp} for ADA2h for the β_4 strand (0.22 ± 0.17) is reasonably

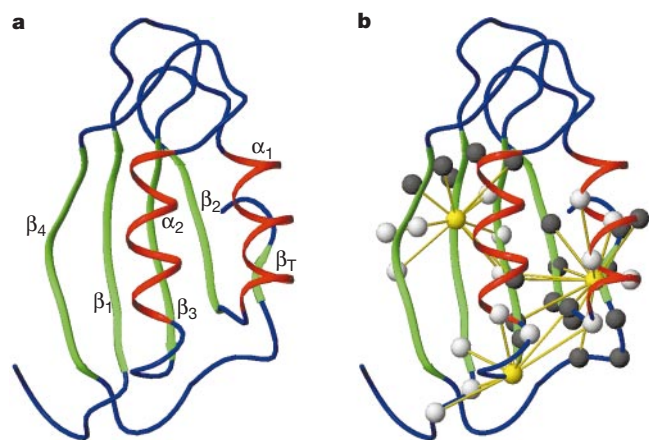


Figure 1 Native structure of acylphosphatase, AcP. **a**, The secondary structure elements are: β_1 (residues 7–13), α_1 (residues 22–33), β_2 (residues 36–42), β_3 (residues 46–53), α_2 (residues 55–56), β_4 (residues 77–85), β_T (residues 93–97); residues between these regions are parts of loops. **b**, the key residues 11, 54 and 94 found from the transition-state analysis are shown as gold spheres on the native structure (see text). They are connected by gold-coloured bonds to the residues (white and black spheres) forming a native contact with them (Y11 has long-range contacts with 47–52 and 78–81, P54 with 5–7 and 34–35, and F94 with 26–31, 36–39 and 50–52). Residues forming the transition-state core identified in the present work are shown as black spheres. They fall into two groups in spatial proximity; there is a larger group (residues 28, 35–39, 51–54 and 90–95) comprising parts of α_1 , β_2 , α_2 and β_T , and a smaller group (residues 11–13, 47 and 78, 79) comprising parts of the β_1 , β_3 and β_4 strands in the native structure.

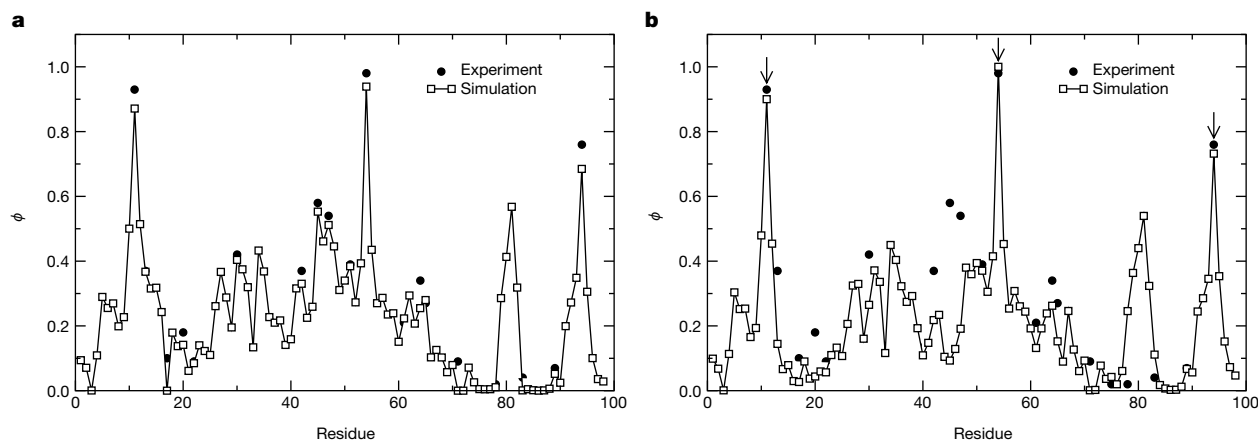


Figure 2 Comparison of ϕ_i^{calc} with ϕ_i^{exp} . **a**, When all 24 of the latter are used as restraints (filled circles); **b**, when only Y11, P54 and F94, are used as restraints (open squares, see text). Only four ϕ_i^{exp} are significantly underestimated; they are V13, T42, G45 and V47. The reason for the underestimation is that this subset of residues forms long-range

contacts mainly to each other (for example, V13 with G45) and are, therefore, almost totally unrestrained in the primitive model when Y11, P54 and F94 are used as restraints; the only common contact is between residues Y11 and V47. Except for V13, which is close to Y11, they are not part of the folding core.

close to the average ϕ_i^{calc} (0.31 ± 0.17) for this region of AcP. Other regions of AcP, where the simulations predict significant non-zero values of ϕ_i^{calc} for which no ϕ_i^{exp} exist, include the hydrophobic residues L6 ($\phi_6^{\text{calc}} = 0.28 \pm 0.10$), A26 ($\phi_{26}^{\text{calc}} = 0.26 \pm 0.16$) and I35 ($\phi_{35}^{\text{calc}} = 0.37 \pm 0.13$); measurements on these residues are in progress (F. Chiti *et al.*, personal communication).

To probe the significance of different residues in determining the structure of TSE, we have performed a series of calculations of the

TSE in which only subsets of the ϕ_i^{exp} are used as restraints. A notable result is that by use of only the three residues with the largest ϕ_i^{exp} (Y11, P54 and F94) as restraints, a correlation coefficient of 0.86 is found between the ϕ_i^{exp} and the ϕ_i^{calc} (Fig. 2b); the cross-validated correlation coefficient (that is, with ϕ_{11}^{calc} , ϕ_{54}^{calc} and ϕ_{94}^{calc} excluded from the comparison) is 0.50. This value is significant in light of the sensitivity of ϕ values to local interactions, which makes their quantitative prediction more difficult than that of the overall fold.

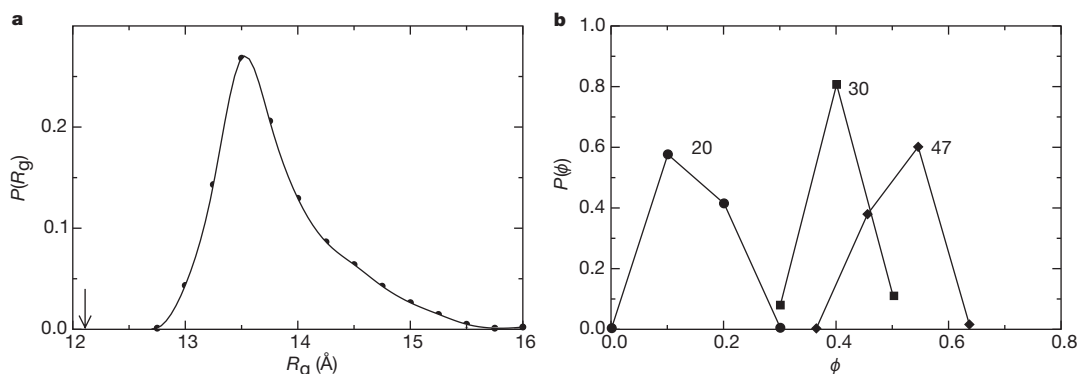


Figure 3 Properties of the transition-state ensemble, TSE, of AcP. **a**, Probability distribution $P(R_g)$ of the radius of gyration for the improved model with $\lambda = 0.85$ and $T = 0.1$ (see Methods). The arrow points to the R_g value of the native state. **b**, Probability $P(\phi)$ of a residue of having a given ϕ value in the TSE. The results for residues V20, A30 and V47 are shown (circles, squares and diamonds, respectively); they have a rather narrow unimodal distribution around the experimental values ($\phi_{20}^{\text{exp}} = 0.18$, $\phi_{30}^{\text{exp}} = 0.42$,

$\phi_{47}^{\text{exp}} = 0.54$). The value $\phi_{47}^{\text{calc}} = 0.56 \pm 0.05$ arises because residue V47 has a contact probability close to 1 with residues 11–14, about 0.5 with residues G15, K40 and N41, and a very low probability of contact with any other residue. $P(\phi)$ remains unimodal for all the residues in the calculation in which only Y11, P54 and F94 are used as a bias in equation (3).

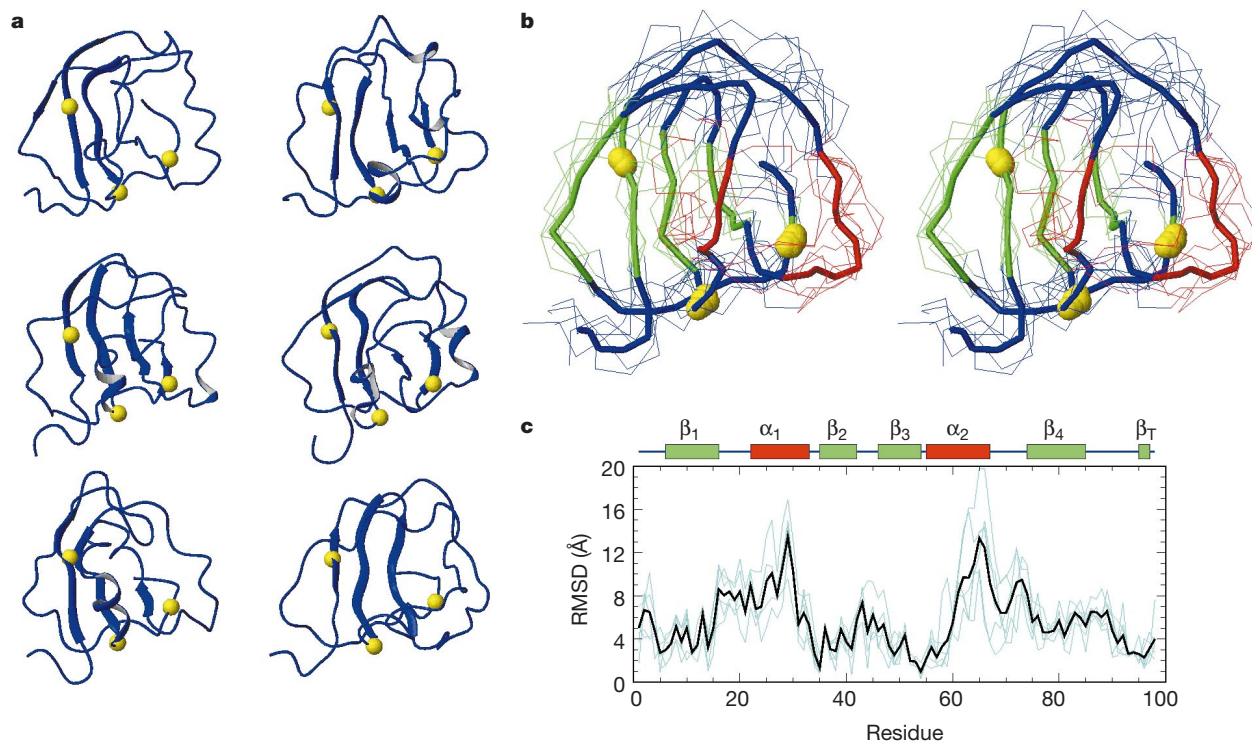


Figure 4 Representation of the TSE of AcP calculated with the improved model (see Methods). **a**, The six most representative conformations of the TSE generated from the structure calculations; more than one-quarter of the 867 generated structures are less than 4 Å in r.m.s. deviation from one of the six structures shown. The three key residues forming the structural core are shown as gold spheres and secondary structural elements are shown as ribbons with arrows indicating the directionality of the β -strands; a method which does not rely on hydrogen-bonding but on comparison with structural templates (M. Schaefer and M.K., unpublished results) was used to identify secondary structures. Preliminary results for molecular dynamics structures obtained with an all-atom potential

and the same ϕ -value restraints indicate that somewhat more secondary structure, particularly for the helices, is present in the TSE (E.P. *et al.*, unpublished results). **b**, The average structure of the transition state (thick lines) derived from the six conformations shown in **a** (thin lines). The residues with a given secondary structure in the native state are indicated with the same colour code as that used in Fig. 1. The average structure is obtained by superimposing the three key residues (gold spheres) for all the structures in the TSE. **c**, Average r.m.s. deviation from the native state as a function of the residue number; the native secondary structural elements are indicated above the diagram.

Nevertheless, the predictions of 21 ϕ_i^{exp} by the 'leave-one-out' method (keeping the three key residues and 20 of the 21 remaining ones) yields a cross-validated correlation coefficient of 0.71. As a further test of the role of the three residues, we compare the ϕ_i^{calc} for all residues obtained with only the three residue restraints and with all the 24 ϕ_i^{exp} restraints; the correlation coefficient is 0.91 (cross-validated 0.83). Other small subsets of residues (not including Y11, P54 and F94) give very poor correlations; the cross-validated correlation coefficients are in the range -0.1 to 0.2. These results show that native-like contacts for the three key residues alone determine the overall fold of the TSE and give information about the additional ϕ values.

Residues Y11, P54 and F94 are part of the hydrophobic core of AcP (ref. 3), and make a large number of contacts with other residues (10, 7 and 14 contacts, respectively). These contacts, which are distributed throughout the structure, create a 'contact network' in the native state (Fig. 1b). There are other residues (for example, K7, V36 and V58) with an even larger numbers of contacts. These contacts, however, are more localized and their ϕ_i^{exp} are smaller, so that they are less important in determining the TSE. We note that the native-like environment for the three key residues can lead to a contact network that determines the fold of most of the protein (see Fig. 1b and legend).

As the primitive model has no stabilizing interactions other than those arising from the ϕ_i^{exp} , the conformational space accessible to the protein is expected to be too large. An 'improved' model (see equation (3) in Methods), in which the energy function is augmented by native-like attractive interactions such that the overall compactness of the TSE is consistent with experimental data on its average solvent exposure, yields a TSE with the same fold as the primitive model. The radius of gyration (R_g) is $13.5 \pm 0.5 \text{ \AA}$ as compared to the native value of $12.6 \text{ \AA}$; the R_g distribution of this TSE is shown in Fig. 3a. The degree of expansion is similar to that observed experimentally for 'molten globule' states in which much of the secondary structure is preserved but most tertiary contacts are not persistent⁹. The core of the transition state of AcP, defined as in Methods, consists of approximately 20 residues including the three key residues and many of their contacts (Fig. 1b).

Figure 4a shows the six most representative structures (that is, the centres of the clusters with most members) of the TSE. Figure 4b shows the average structure obtained by superposing all the members of the TSE; also shown are the structures corresponding to the dominant clusters (those in Fig. 4a), which yield the same average structure. Comparison of the average structure in Fig. 4b with Fig. 1a demonstrates that the fold of the TSE is native-like, but that it is significantly expanded relative to the native structure. There is a marked tendency for the β -sheets to be present, particularly β_2 and β_3 , whereas the location of the helical regions is less well defined. Calculations incorporating additional local restraints, to account for the recent result of Taddei *et al.*¹⁰ that stabilizing helix 2 increases the folding rate, suggest that this helix is in fact present in the TSE but that it is highly disordered relative to the rest of the structure.

The r.m.s. deviation of the TSE from the native state as a function of residue number is shown in Fig. 4c; it makes clear how much the average structure deviates from the native state (some parts by more than $15 \text{ \AA}$), even though the fold is preserved. There are some non-native interactions (about 30% of the contacts are non-native). Overall the contact order¹¹ in the TSE (13.7) is significantly lower than in the native state (20.8), indicating that only about 70% of the native-like long-range contacts are formed. The distributions of the fraction of native contacts in the TSE obtained for AcP are relatively narrow; that is, the same residues make contact with each other in most members of the TSE (Fig. 3b). This finding for AcP supports the conclusions of Fersht *et al.*¹² for chymotrypsin inhibitor 2 that fractional ϕ values arise from the formation of an incomplete set of native contacts in all the members of the TSE rather than from the

alternative possibility that some molecules in the TSE have most of their native contacts and others have only a few or none.

The present analysis introduces an approach for generating the molecular structure of the TSE from experimental ϕ values. It provides a check on the consistency of the experimental data, and determines the importance of specific structural elements and inter-residue interactions. If other parameters concerning the TSE are available from experiment, the present approach can be refined by including them in the energy function used for the structure determination. As the TSE is based on experiment, it complements existing theoretical studies^{13–22}. The demonstration for AcP that the contacts of a small number of key residues determines the native-like fold provides a microscopic characterization of the nucleation step of the folding process². These key residues form an extensive 'contact network' of native-like interactions in the TSE. This explains the role of these residues, which are not in direct contact with each other, as crucial elements in determining the transition state. The global contact network leads to a rather narrow TSE, whose core elements correspond to an expanded form of the native structure. We anticipate that the application of this method of analysis to a range of proteins will lead to a clearer understanding of the nature of the folding transition state and its relationship to the native structure. □

Methods

Definition of ϕ_i^{calc}

We have adopted a definition of ϕ_i^{calc} similar to one used previously in molecular-dynamics simulations^{14,23}. For residue i ,

$$\phi_i^{\text{calc}} = \frac{\langle N_i \rangle_{\text{TSE}}}{N_i^N} \quad (1)$$

where N_i is the number of native contacts formed by residue i ; the average in the numerator, $\langle N_i \rangle_{\text{TSE}}$, is over all the conformations of the TSE, and in the denominator, N_i^N refers to the native state structure. Contacts are defined to exist when the α positions are closer than $8.5 \text{ \AA}$ (ref. 24) (see below) and the residues are more than two neighbours apart along the chain (that is, contacts $i+1$ and $i+2$ are omitted). In equation (1) all native contacts give an equal contribution whereas non-native contacts are not included.

Models

Each amino acid is represented as a sphere of radius $4.25 \text{ \AA}$ with an inner hard-core radius of $2.35 \text{ \AA}$ to take into account excluded volume effects. Each sphere is centred at the position of the α atom and is connected by rigid bonds of length $3.8 \text{ \AA}$ to its neighbours along the chain²⁵. Thus, two residues are in contact when they are closer than $8.5 \text{ \AA}$, and non-neighbouring amino acids are not permitted to be closer than $4.7 \text{ \AA}$. Comparison of the contacts derived from the present α model with those obtained from an all-atom model shows good agreement throughout the protein.

In the minimum interaction model ('primitive model') the energy of a given conformation, represented by its contact map S , is defined as:

$$E_{\text{min}}(S) = \frac{N^N}{N_\phi} \sum_i [\phi_i^{\text{exp}} - \phi_i^{\text{calc}}(S)]^2 \quad (2)$$

that is, the only attractive energy term between non-bonded residues is from the values of ϕ_i^{exp} ; N_ϕ^N is the total number of contacts in S^N , and N_ϕ is the number of available ϕ_i^{exp} . In the more realistic interaction model (the 'improved model'), an attractive (Go-like²⁶) interaction term for the native state contacts is introduced. It has the form:

$$E_\lambda(S) = \lambda E_{\text{min}}(S) + (1 - \lambda) E_{\text{Go}}(S) \quad (3)$$

where the Gō-model energy, $E_{\text{Go}}(S)$, is

$$E_{\text{Go}}(S) = -\epsilon_1 \sum_{ij} S_{ij} S_{ij}^N + \epsilon_2 \sum_{ij} S_{ij} (1 - S_{ij}^N) \quad (4)$$

where S^N is the contact map of the native state. The element S_{ij} of a contact map is defined to be 1 if residues i and j are in contact and 0 otherwise. The phase diagram of E_{Go} is well known²⁶ and the two parameters ϵ_1 and ϵ_2 are fixed at $\epsilon_1 = 1$ and $\epsilon_2 = 0.1$ (M.V., E.P., C.M.D. and M.K., unpublished calculations). The parameter λ tunes the compactness of the TSE. For $\lambda = 0$ (the standard Gō-model) and a temperature $T = 0.1$ in arbitrary units (see below), we obtain a native-like value for the radius of gyration, $R_g = 12.7 \pm 0.1 \text{ \AA}$, and for $\lambda = 1$ (the 'primitive model') we obtain $R_g = 15.2 \pm 0.6 \text{ \AA}$. To choose a value for λ we performed Monte Carlo sampling at $T = 0.1$ to generate the TSE for several values of λ . For each value of λ , we computed the average accessible surface area (ASA) for the

conformations in the TSE using a probe of 3.1 Å that matches the standard one of 1.4 Å from all-atom calculations.

Using experimentally derived m values we estimate ASA_{TSE}^{exp} by using the empirical formula^{14,27}

$$RT \frac{m_u}{m} = \frac{ASA_{TSE}^{exp} - ASA_{NSE}}{ASA_{RCE} - ASA_{NSE}} \quad (5)$$

where RCE and NSE stand for random-coil ensemble and native-state ensemble, respectively; m and m_u represent the derivative with respect to the denaturant concentration of the equilibrium constant and the rate of unfolding, respectively. The measured ratio is 0.23 for AcP (ref. 3). The intersection between the curves $ASA_{TSE}(\lambda)$ and ASA_{TSE}^{exp} gives the selected value, $\lambda = 0.85$. Use of only E_{Go} (equation (4)) in the Monte Carlo sampling yields a correlation coefficient of 0.09 between ϕ_i^{calc} and ϕ_i^{exp} ; that is, there is essentially no correlation with the TSE.

Given the empirical energy function (equations (2)–(4)), a Monte Carlo method is used to determine the ensemble of structures, TSE, making up the transition state. Monte Carlo moves are bond-length-conserving crankshaft rotations that are restricted by steric restraints among neighbouring residues along the sequence and accepted according to the standard Metropolis criterion at the Monte Carlo temperature T (ref. 24). The choice of $T = 0.1$ is made by requiring that both for $\lambda = 0$ and for $\lambda = 1$ the protein is in a collapsed state.

Clustering

We assume that stable cores are formed by residues whose relative fluctuations are small for the given energy function. The 'distance' between residues used in clustering is related to the size of the fluctuations in their distance. Specifically, the distance f_{ij} between residues i and j is defined as a 'fluctuation ratio':

$$f_{ij} = \frac{\langle u_{ij}^2 \rangle}{\langle r_{ij}^2 \rangle} \quad (6)$$

where the averages are taken over the conformations in the TSE, $u_{ij} = r_{ij} - \langle r_{ij} \rangle$, and r_{ij} are the cartesian distances between residues i and j in a given conformation in the TSE. The fluctuation ratio is related to the Berry parameter²⁸ used in the description of systems without a reference state when the Lindemann criterion²⁹ is unsuitable. The clustering using the distances f_{ij} was performed by using the SPC algorithm³⁰, and the resulting dendrograms were analysed with an algorithm provided by G. Getz and E. Domany (personal communication). The average fluctuation ratio over all pairs in the core is 0.019 and that over the entire protein is 0.049, more than twice as large; for comparison, the average fluctuation ratio in the native state with the energy function given in equation (4) is 0.010.

Received 8 June; accepted 8 December 2000.

- Matouschek, A., Kellis, J. T., Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122–126 (1989).
- Fersht, A. R. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York, 1999).
- Chiti, F. *et al.* Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**, 1005–1009 (1999).
- Polanyi, J. C. & Zewail, A. H. Direct observation of the transition state. *Acc. Chem. Res.* **28**, 119–132 (1995).
- Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M. & Smith, M. Redesigning enzyme structure by site-directed mutagenesis. Tyrosyl tRNA ATP binding. *Nature* **299**, 756–758 (1982).
- Fersht, A. R., Leatherbarrow, R. J. & Wells, T. N. Structure-activity relationships in engineered proteins. Analysis of use of binding energy by linear free energy relationships. *Biochemistry* **26**, 6030–6038 (1987).
- Wüthrich, K. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* **243**, 45–50 (1989).
- Villegas, V., Martínez, J. C., Avilés, F. X. & Serrano, L. Structure of the transition state in the folding

- process of human procarboxypeptidase A2 activation domain. *J. Mol. Biol.* **283**, 1027–1036 (1998).
- Kuwajima, K. The molten globule state of α -lactalbumin. *FASEB J.* **10**, 102–109 (1996).
- Taddei, N. *et al.* Stabilisation of α -helices by site-directed mutagenesis reveals the secondary structure in the transition state for acylphosphatase folding. *J. Mol. Biol.* **300**, 633–647 (2000).
- Plaxco, K. W., Simons, K. & Baker, D. Contact order, transition state placement and the refolding rate of single domain proteins. *J. Mol. Biol.* **277**, 985–994 (1998).
- Fersht, A. R., Itzhaki, L. S., elMasry, N. F., Matthews, J. M. & Otzen, D. E. Single versus parallel pathways of protein folding and fractional structure in the transition state. *Proc. Natl Acad. Sci. USA* **91**, 10426–10429 (1994).
- Daggett, V., Li, A., Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* **257**, 430–440 (1996).
- Lazaridis, T. & Karplus, M. "New View" of protein folding reconciled with the old through multiple unfolding simulations. *Science* **278**, 1928–1931 (1997).
- Boczek, E. M. & Brooks, C. L. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science* **269**, 393–396 (1995).
- Muñoz, V. & Eaton, W. A. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl Acad. Sci. USA* **96**, 11311–11316 (1999).
- Galitskaya, O. V. & Finkelstein, A. V. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl Acad. Sci. USA* **96**, 11299–11304 (1999).
- Alm, E. & Baker, D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl Acad. Sci. USA* **96**, 11305–11310 (1999).
- Micheletti, C., Banavar, J. R., Maritan, A. & Seno, F. Protein structures and optimal folding from a geometrical variational principle. *Phys. Rev. Lett.* **82**, 3372–3376 (1999).
- Shoemaker, B. A., Wang, J. & Wolynes, P. G. Exploring structures in protein folding funnels with free energy functionals: The transition state ensemble. *J. Mol. Biol.* **287**, 675–694 (1999).
- Clementi, C., Jennings, P. A. & Onuchic, J. N. How native-state topology affects the folding of dihydrofolate reductase. *Proc. Natl Acad. Sci. USA* **97**, 5871–5876 (2000).
- Li, L., Mirny, L. A. & Shakhnovich, E. I. Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. *Nature Struct. Biol.* **7**, 336–342 (2000).
- Li, A. & Daggett, V. Characterization of the transition state of protein unfolding by use of molecular dynamics: Chymotrypsin inhibitor 2. *Proc. Natl Acad. Sci. USA* **91**, 10430–10434 (1994).
- Vendruscolo, M. & Domany, E. Pairwise contact potentials are unsuitable for protein folding. *J. Chem. Phys.* **109**, 11101–11108 (1998).
- Vendruscolo, M. & Domany, E. Efficient dynamics in the space of contact maps. *Fold. Des.* **3**, 329–336 (1998).
- Zhou, Y. & Karplus, M. Interpreting the folding kinetics of helical proteins. *Nature* **401**, 400–403 (1999).
- Tanford, C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* **24**, 1–95 (1970).
- Berry, R. S., Beck, T. L., Davis, H. L. & Jellinek, J. In *Advances in Chemical Physics* Vol 70, (eds Prigogine, I. & Rice, S. A.) 75–135 (Wiley, New York, 1988).
- Zhou, Y., Vitkup, D. & Karplus, M. Native proteins are surface-molten solids: Application of the Lindemann criterion for the solid versus liquid state. *J. Mol. Biol.* **285**, 1371–1375 (1999).
- Domany, E. Superparamagnetic clustering of data. The definitive solution of an ill posed problem. *Physica A* **263**, 158–169 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank F. Chiti and A. Fersht for discussions and comments on this work. We also thank G. Ramponi and N. Taddei for continuing collaborations involving AcP. The Oxford Centre for Molecular Science is supported by BBSRC, EPSRC and MRC. M.V. is supported by an EMBP long-term fellowship. E.P. is supported in Oxford by an EC fellowship. C.M.D. is supported in part by a programme grant from the Wellcome Trust. M.K. Thanks Oxford University for inviting him to spend a year as Eastman Visiting Professor. Much of this work was done while he was in Oxford; the part done at Harvard was supported in part by the National Institutes of Health.

Correspondence and requests for materials should be addressed to C.M.D. (e-mail: chris.dobson@chem.ox.ac.uk) or M.K. (e-mail: marci@tammy.harvard.edu).

A Toll-like receptor recognizes bacterial DNA

Hiroaki Hemmi, Osamu Takeuchi, Taro Kawai, Tsuneyasu Kaisho, Shintaro Sato, Hideki Sanjo, Makoto Matsumoto, Katsuaki Hoshino, Hermann Wagner, Kiyoshi Takeda & Shizuo Akira

Nature **408**, 740–744 (2000).

Figure 1 was printed incorrectly in this paper owing to a font error. The correct Fig. 1 is reproduced below. □

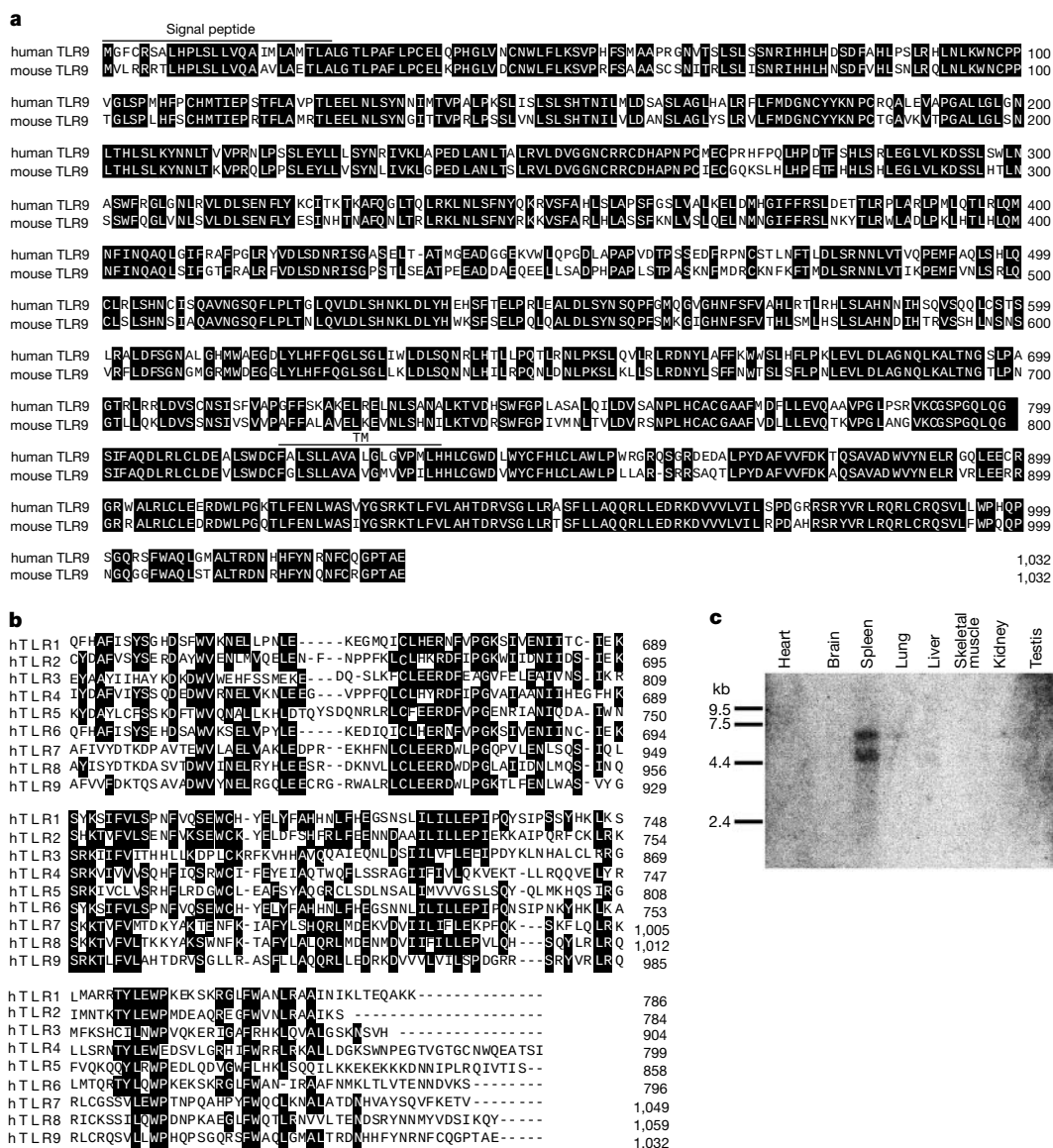


Figure 1 Amino-acid sequences and tissue distribution of TLR9. **a**, Alignment of human and mouse TLR9. Identical amino acids are indicated by a solid box. Human and mouse TLR9 share an overall amino-acid identity of 75.5%. The predicted signal peptide (human and mouse, residues 1–25) and transmembrane segments (TM; human, 819–836; mouse, 820–837) are indicated. During the preparation of this manuscript, the

sequences for human TLR9 were deposited in GenBank by two other groups (accession numbers NM017442 and AF245704). **b**, Alignment of the cytoplasmic domains of human TLR family members. Amino-acid residues conserved in least four molecules are highlighted by solid boxes. **c**, A mouse multiple-tissue northern blot (Clontech) containing 2 μ g of poly(A)⁺ RNA was probed with a mouse TLR9 cDNA fragment.

The subtleties of spin

A page of the latest centrifuges and centrifuge accessories.



Nalge's centrifuge bottles, good for 27,500g.

Centrifuge bottles

Nalge Nunc www.nalgenunc.com

Leak-proof at maximum speed

Transparent and autoclavable, Nalge centrifuge bottles are built to be leak-proof at rated speeds of up to 27,500g on hazardous materials when used with polypropylene sealing caps. The bottles are manufactured from rigid polycarbonate and are also available in a polypropylene copolymer version. Choose from nominal capacities of 250 ml, 500 ml and 1,000 ml.

Reader Service No. 100

Napco 2028R

Precision/Napco www.PrecisionSci.com

Refrigerated centrifuge for cell culture work

The Napco 2028R provides both high- and low-speed centrifugation to handle a wide variety of applications, including cell harvesting and culture refreshment. With a 4 × 280-ml swinging bucket rotor and an assortment of fixed-angle rotors, it offers capacities of twenty 15-ml conical tubes at 3,000g, or six 50-ml conical tubes at over 12,500g, and comes with sealed microtitre plate carriers and cyto-



Harvest home: the Napco 2028R, for cell cultures.

centrifugation accessories. Rotor exchange is completed in less than five seconds via a device that allows rotors to glide into position on the drive shaft and lock automatically, without the use of traditional nuts and wrenches. CFC-free refrigeration protects heat-labile samples, and variable acceleration rates preserve fragile samples and maximize separation quality.

Reader Service No. 101

Sigma 2-5

Philip Harris Scientific www.philipharris.co.uk

Compact and quiet

Designed for day-to-day use in small medical laboratories, the Sigma 2-5 general-purpose centrifuge has a maximum speed of 3,900 r.p.m. (2,310g). Low-speed operation from 100 r.p.m. is also possible. It can accommodate four 100-ml or six microplates, while a four-place swing-out rotor allows 16 × 15-ml, 16 × 7-ml vacutainers and four universal falcon tubes. A sealed bucket is available for use with 15-ml and 50-ml falcon tubes. New rotors include a fixed-angle 20 × 15-ml falcon tube rotor and a two-place rotor for up to six microplates.

Reader Service No. 102

Avanti J-20XP

Beckman Coulter www.beckmancoulter.com

High throughput with no loss of g-force

This centrifuge combines high performance — 26,000 r.p.m. and 82,000g — with a capacity of six litres, which it will spin in under 10 minutes. A high-torque, brushless drive shortens acceleration and deceleration times. The J-20XP accommodates all Avanti J rotors and labware, including one-litre bottles, at forces of up to 15,900g, and 50-ml tubes at 82,000g. The rotors feature labware with a wide mouth to facilitate sample handling, a simplified bottle closure and a contoured spatula for efficient sample removal.

Reader Service No. 103

MiniSpin and MiniSpin Plus

Eppendorf www.eppendorf.com

Dynamic duo

MiniSpin microcentrifuges feature rapid acceleration, a maximum braking time of 13 seconds, low noise levels, and a footprint smaller than a sheet of A4 paper. During routine runs, such as pelleting precipitated DNA, sample heating is kept to a bare minimum. When loaded with 1.5-ml or 2.0-ml Eppendorf tubes, the MiniSpin can attain a maximum speed of 13,400 r.p.m., or an r.c.f. of 12,100g. The MiniSpin Plus is capable of 14,000 r.p.m.



Small is beautiful: Eppendorf's MiniSpins.

(14,000g) and can be set to either short-spin or continuous operation, which may be required if, for example, complex research tasks need to be completed in a cold room.

Reader Service No. 104

Sorvall Colortone Superspeed rotors

Kendro www.kendro.com

Rotor identification at a glance

These rotors, for use with the Sorvall RC-5 series, RC-26 Plus and Super T 21 centrifuges, are available in anodized blue, red, green, purple, gold and silver. Users can designate different colours for biohazardous or radioactive applications. Colour coding by protocol or contamination-susceptible sample type can help to prevent cross-contamination, and ownership indicated by assigning different colours to different laboratories, research groups, or centrifuges.

Reader Service No. 105

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENT

BACs • YACs • PACs cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse and rat libraries as well as a custom screening service for the libraries.

Research Genetics

U.S. or Canada 800-533-4363

U.K. 0-800-89-1393

FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

READER ENQUIRY NO. 6